THE UNIVERSITY OF WARWICK

Library Declaration and Deposit Agreement

1. STUDENT DETAILS

Please complete the following:

Full name: PHILLIP PETCHER

University ID number: 0114251

2. THESIS DEPOSIT

2.1 I understand that under my registration at the University, I am required to deposit my thesis with the University in BOTH hard copy and in digital format. The digital version should normally be saved as a single pdf file.

2.2 The hard copy will be housed in the University Library. The digital version will be deposited in the University's Institutional Repository (WRAP). Unless otherwise indicated (see 2.3 below) this will be made openly accessible on the Internet and will be supplied to the British Library to be made available online via its Electronic Theses Online Service (EThOS) service.
[At present, theses submitted for a Master's degree by Research (MA, MSc, LLM, MS or MMedSci) are not being deposited in WRAP and not being made available via EthOS. This may change in future.]

2.3 In exceptional circumstances, the Chair of the Board of Graduate Studies may grant permission for an embargo to be placed on public access to the hard copy thesis for a limited period. It is also possible to apply separately for an embargo on the digital version. (Further information is available in the *Guide to Examinations for Higher Degrees by Research*.)

2.4 *If you are depositing a thesis for a Master's degree by Research, please complete section (a) below. For all other research degrees, please complete both sections (a) and (b) below:*

(a) Hard Copy

I hereby deposit a hard copy of my thesis in the University Library to be made publicly available to readers ~~(please delete as appropriate) EITHER~~ immediately ~~OR after an embargo period of ..... months/years as agreed by the Chair of the Board of Graduate Studies~~.

I agree that my thesis may be photocopied. YES / ~~NO~~ (*Please delete as appropriate*)

(b) Digital Copy

I hereby deposit a digital copy of my thesis to be held in WRAP and made available via EThOS.

Please choose one of the following options:

My thesis can be made publicly available online. YES / ~~NO~~ (*Please delete as appropriate*)

3. GRANTING OF NON-EXCLUSIVE RIGHTS

Whether I deposit my Work personally or through an assistant or other agent, I agree to the following:

Rights granted to the University of Warwick and the British Library and the user of the thesis through this agreement are non-exclusive. I retain all rights in the thesis in its present version or future versions. I agree that the institutional repository administrators and the British Library or their agents may, without changing content, digitise and migrate the thesis to any medium or format for the purpose of future preservation and accessibility.

4. DECLARATIONS

(a) I DECLARE THAT:

- I am the author and owner of the copyright in the thesis and/or I have the authority of the authors and owners of the copyright in the thesis to make this agreement. Reproduction of any part of this thesis for teaching or in academic or other forms of publication is subject to the normal limitations on the use of copyrighted materials and to the proper and full acknowledgement of its source.

- The digital version of the thesis I am supplying is the same version as the final, hard-bound copy submitted in completion of my degree, once any minor corrections have been completed.

- I have exercised reasonable care to ensure that the thesis is original, and does not to the best of my knowledge break any UK law or other Intellectual Property Right, or contain any confidential material.

- I understand that, through the medium of the Internet, files will be available to automated agents, and may be searched and copied by, for example, text mining and plagiarism detection software.

(b) IF I HAVE AGREED (in Section 2 above) TO MAKE MY THESIS PUBLICLY AVAILABLE DIGITALLY, I ALSO DECLARE THAT:

- I grant the University of Warwick and the British Library a licence to make available on the Internet the thesis in digitised format through the Institutional Repository and through the British Library via the EThOS service.

- If my thesis does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my thesis available in digital format and that this permission encompasses the rights that I have granted to the University of Warwick and to the British Library.

5. LEGAL INFRINGEMENTS

I understand that neither the University of Warwick nor the British Library have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights, breach of contract or of any other right, in the thesis.

Please sign this agreement and return it to the Graduate School Office when you submit your thesis.

Student's signature: ..... Date: 27/09/2011

Time of Flight Diffraction and Imaging (TOFDI)

Phillip Anthony Petcher

THE UNIVERSITY OF
WARWICK

A thesis submitted for the degree of

Doctor of Philosophy in Physics

University of Warwick, Department of Physics

September 2011

Contents

List of Figures

List of Tables

Acknowledgements

This work is based upon an original idea by Steve Dixon. Not only has Steve provided vital technical input to this project, he has supported my development as a researcher. Thank you to Steve for getting me through my PhD, and helping me in my career beyond it.

Thank you to my family who have always supported me, my dad Anthony, my mum Laraine, and my big sister Karolyn. Without them, I couldn't have got anywhere. My dad has always encouraged me to do my best, and my mum has made sure I didn't take things too seriously. My sister taught me how to tie my shoelaces and helped me improve my maths when I was eight, and I think that set me on the right path.

Thank you to my colleagues, who are all brilliant people. José Francisco Hernández Valle is possibly the nicest man in the world, his dedication to his work and his kindness is inspirational. Thank you for helping me out when I needed it Francisco, it was very much appreciated. Thanks to Timothy Harrison, I enjoyed our discussions on new technology, and there are journal articles with my name on because you asked me for help. Iain Baillie helped me to get industrial funding, and linked my research to the real world, both of which made my PhD better.

Thank you to the University of Warwick, the Engineering and Physical Sciences Research Council (EPSRC), and Corus, for financially supporting the project.

Thank you to anybody I have missed for being great as well, and not minding that your name has been omitted. Your support has been appreciated.

Declaration

I declare that the work presented in this thesis is my own except where stated otherwise, and was carried out entirely at the University of Warwick during the period between October 2007 and September 2011, under the supervision of Prof. Steve Dixon. The research reported here has not been submitted, either wholly or in part, in this or any other academic institution, for admission to a higher degree. Parts of this work, and other work not reported in this thesis, have appeared in other forms, listed here.

Peer-reviewed journal articles:

- P. A. Petcher and S. Dixon, "Parabola detection using matched filtering for ultrasound B-scans," *Ultrasonics*, vol. 52, no. 1, pp. 138-144, Jan. 2012.
- P. A. Petcher and S. Dixon, "Non-contact position tracking using phase correlation of sample images for NDT applications," *Insight*, vol. 53, no. 4, pp. 185-191, Apr. 2011.
- P. A. Petcher and S. Dixon, "A modified Hough transform for removal of direct and reflected surface waves from B-scans," *NDT & E International*, vol. 44, no. 2, pp. 139-144, Mar. 2011.
- S. Dixon, T. J. Harrison, and P. A. Petcher, "Phase changes of ultrasonic bulk waves through focusing measured using a noncontact ultrasonic method," *Applied Physics Letters*, vol. 97, no. 5, p. 054101, Aug. 2010.

Conference proceedings:

- P. A. Petcher and S. Dixon, "Time of flight diffraction and imaging (TOFDI) combining B-scans and cross-sectional imaging," in *AIP Conference Proceedings*, vol. 1335, San Diego (California), USA: AIP, 2011, pp. 744-751.
- P. A. Petcher and S. Dixon, "Use of a sparse array for time of flight diffraction and imaging (TOFDI)," in *Proceedings of the National Seminar & Exhibition on Non-Destructive Evaluation*, Tiruchirapalli, India: NDT.net, 2010, pp. 309-314.
- P. A. Petcher and S. Dixon, "Wideband ultrasonic time of flight diffraction combining B-scans and cross-sectional imaging," in *AIP Conference Proceedings*, vol. 1096, Chicago (Illinois), USA: AIP, 2009, pp. 650-657.

Presentations at national and international conferences:

- P. A. Petcher and S. Dixon, "Time of flight diffraction and imaging (TOFDI) combining B-scans and cross-sectional imaging," presented at the *49th Annual British Conference on Non-Destructive Testing (NDT 2010)*, Cardiff, UK, Sep. 2010. Talk.

- P. A. Petcher and S. Dixon, "Wideband ultrasonic time of flight diffraction combining B-scans and cross-sectional imaging," presented at the *37th Annual Review of Progress in Quantitative Nondestructive Evaluation (QNDE 2010)*, San Diego (California), USA, July 2010. Poster.
- P. A. Petcher and S. Dixon, "Use of a sparse array for time of flight diffraction and imaging (TOFDI)," presented at the *National Seminar on NDE (NDE 2009)*, Tiruchirapalli, India, Dec. 2009. Talk (delivered by S. Dixon).
- P. A. Petcher and S. Dixon, "Developments in time of flight diffraction and imaging (TOFDI)," presented at the *48th Annual British Conference on Non-Destructive Testing (NDT 2009)*, Blackpool, UK, Sep. 2009. Talk.
- P. A. Petcher and S. Dixon, "Wideband ultrasonic time-of-flight diffraction combining B-scans and cross-sectional imaging," presented at the *47th Annual British Conference on Non-Destructive Testing (NDT 2008)*, Cheshire, UK, Sep. 2008. Talk.
- P. A. Petcher and S. Dixon, "Wideband ultrasonic time of flight diffraction combining B-scans and cross-sectional imaging," presented at the *35th Annual Review of Progress in Quantitative Nondestructive Evaluation (QNDE 2008)*, Chicago (Illinois), USA, July 2008. Poster.

Presentations at other events:

- P. A. Petcher and S. Dixon, "Defective Metal, Effective Detection!," presented at the *Annual Postgraduate Research Poster Competition*, Nottingham, UK: Nottingham Trent University, July 2010. Poster.
- P. A. Petcher and S. Dixon, "Defective Metal, Effective Detection!," presented at the *5th Annual Postgraduate Research Poster Competition*, Coventry, UK: University of Warwick, May 2010. Poster.
- P. A. Petcher and S. Dixon, "Defective Metal, Effective Detection!," presented at the *Annual Postgraduate Research Poster Competition*, Nottingham, UK: University of Nottingham, July 2009. Poster.
- P. A. Petcher and S. Dixon, "Defective Metal, Effective Detection!," presented at the *4th Annual Postgraduate Research Poster Competition*, Coventry, UK: University of Warwick, May 2009. Poster.
- P. A. Petcher and S. Dixon, "Crouching Transducer, Hidden Defect," presented at the *Annual Postgraduate Research Poster Competition*, Keele, UK: Keele University, July 2008. Poster.
- P. A. Petcher and S. Dixon, "Crouching Transducer, Hidden Defect," presented at the *3rd Annual Postgraduate Research Poster Competition*, Coventry, UK: University of Warwick, May 2008. Poster.

Abstract

Time of flight diffraction and imaging (TOFDI) is based on time of flight diffraction (TOFD), adding cross-sectional imaging of the sample bulk by exploiting the scattering of ultrasonic waves from bulk defects in metals. Multiple wave modes are emitted by a pulsed laser ultrasound ablative source, and received by a sparse array of receiving electromagnetic acoustic transducers (EMATs), for non-contact (linear) scanning, with mode-conversions whenever waves are scattered.

Standard signal processing techniques, such as band-pass filters, reduce noise. A B-scan is formed from multiple data captures (A-scans), with time and scan position axes, and colour representing amplitude or magnitude. B-scans may contain horizontal lines from surface waves propagating directly from emitter to receiver, or via a back-wall, and angled lines after reflection off a surface edge. A Hough transform (HT), modified to deal with the constraints of a B-scan, can remove such lines. A parabola matched filter has been developed that identifies the features in the B-scan caused by scattering from point-like defects, reducing them to peaks and minimising noise. Multiple B-scans are combined to reduce noise further. The B-scan is also processed to form a cross-sectional image, enabling detection and positioning of multiple defects.

The standard phase correlation technique applied to camera images, has been used to track the relative position between transducer and sample. Movement has been determined to sub-pixel precision, with a median accuracy of 0.01mm of linear movement (0.06 of a pixel), despite uneven illumination and the use of a basic low resolution camera.

The prototype application is testing rough steel products formed by continuous casting, but the techniques created to facilitate operation of TOFDI are applicable elsewhere.

Abbreviations

AC	Alternating Current
ADC	Analogue-to-Digital Converter
AIP	American Institute of Physics
AWGN	Additive White Gaussian Noise
BNC	Bayonet Neill-Concelman
BS	British Standards
CPU	Central Processing Unit
DAC	Distance Amplitude Correction
DC	Direct Current
DFT	Discrete Fourier Transform
EBSD	Electron BackScatter Diffraction
EMAT	ElectroMagnetic Acoustic Transducer
EMF	ElectroMotive Force
EMI	ElectroMagnetic Interference
EPSRC	Engineering and Physical Sciences Research Council
FEA	Finite Element Analysis
FEM	Finite Element Method
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FMC	Full Matrix Capture
FT	Fourier Transform
FT-SAFT	Fourier Transform Synthetic Aperture Focusing Technique
FWHM	Full Width at Half Maximum
HT	Hough Transform
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
IFT	Inverse Fourier Transform
IHT	Inverse Hough Transform
MATLAB	MATrix LABoratory
Nd:YAG	Neodymium-doped Yttrium Aluminium Garnet ($Y_3Al_5O_{12}$)
NDE	NonDestructive Examination
NDT	Non-Destructive Testing
NPD	Normal Probe Diffraction
PC	Personal Computer

PCI	Peripheral Component Interconnect
PSF	Point Spread Function
PWR	Pressurised Water Reactor
QNDE	Quantitative NonDestructive Evaluation
RAM	Random-Access Memory
RF	Radio Frequency
SAFT	Synthetic Aperture Focusing Technique
SAW	Surface Acoustic Wave
SH	Shear Horizontal
SI	Système International d'unités (International System of Units)
SIR	Spatial Impulse Response
SNR	Signal-to-Noise Ratio
SV	Shear Vertical
TFM	Total Focusing Method
TOFD	Time Of Flight Diffraction
TOFDI	Time Of Flight Diffraction and Imaging
UK	United Kingdom
USA	United States of America
USB	Universal Serial Bus
VTFM	Vector Total Focusing Method

Chapter 1

Introduction

This chapter introduces the basic concepts of non-destructive testing, and the requirements that industry expects such testing to meet. Time of flight diffraction and imaging is introduced as a non-destructive testing technique that meets these requirements, and is put into context regarding the standard time of flight based technique, time of flight diffraction, and newer techniques using time of flight, including those performing cross-sectional imaging.

1.1 Non-destructive testing

Components usually fail due to defects such as a crack causing a reduction in strength, and such defects may arise during manufacture [1] or service of the product [2]. A prototype component may be tested destructively, by placing the component under adverse conditions and subsequently observing its condition. For example, motorcycle helmets can be tested by placing them in a simulated crash situation, but this leaves the helmet no longer fit for purpose. Destructive testing always leaves the sample unfit for its purpose, which is not only expensive, it clearly cannot be done for every product produced as this would result in zero surviving products. Consequently, only a small sample of a given component type can be tested destructively, and regarding a component currently in service, testing destructively could only be justified if there were a very large number of identical components and the condition of all of them was expected to be very similar to those tested destructively. The samples chosen would then of course not be synchronised to the rest of the components with respect to their condition, as they would need to be replaced.

Non-destructive testing (NDT) is testing that does not destroy the component, hence there should be no change in the properties important to its use [3]. NDT can mean that the sample is completely unaltered, or it can mean that only damage is caused which does not stop the sample being used as intended. For example, a steel billet that is the output of a continuous casting procedure may need testing for major internal defects. Minor surface damage caused by the use of ablative laser ultrasound is inconsequential [4], and hence the testing can be considered non-destructive. However, such surface damage would not be allowed if testing a human subject, and in this case, the procedure would have to do no damage at all to be considered non-destructive.

This work concerns itself with ultrasonic NDT, and the aim has been to create a system to provide an image of internal defects within bulk metals, particularly steel. Assumptions made are for a small number of relatively strong scatterers in a conducting homogeneous isotropic bulk medium.

1.2 Requirements and objectives

Whilst off-line testing is adequate for many products and components, continuous on-line monitoring, such as for continuously cast steel, can reduce wastage by detecting faults early. Industry requires the timely detection, identification, positioning, and sizing, of defects present in bulk samples that vary significantly in geometry and composition; additional common constraints are a hostile environment, such as a hot and rough moving sample with slight variations in basic geometry between samples in a given set [5]. Contact transducers may not be able to work on high temperature surfaces, and the high speed of the sample may make it too difficult to maintain consistent coupling [6]; non-contact transducers remove variation caused by a couplant layer [7]. The move to automated inspections is due to the team-to-team variability often observed [8], and of course for reasons of speed and cost. Automation can remove variability to a large extent, or at least replace it with a consistent level of variability based upon the specifications of the hardware and software used for the inspection.

This project aims to meet these requirements using a new technique known as time of flight diffraction and imaging (TOFDI), essentially wideband ultrasonic time of flight diffraction (TOFD) combining B-Scans and cross-sectional imaging. The approach fuses conventional TOFD techniques with new signal and image processing methods, which extract the maximum information from a B-scan. The first step is to increase the potentially available information using an array of discrete transducers receiving ultrasound. Typically these will be non-contact electromagnetic acoustic transducers (EMATs) due to the requirement for survival in hot and rough conditions, rather than a laser interferometer which has a relatively high price, or a piezoelectric probe which requires a couplant and would be unable to deal with hot and rough conditions whilst scanning. A sparse discrete array is better able to deal with complex geometry situations due to greater flexibility in positioning individual elements than a phased array or other dense array types. In addition, it does not require all the transducers to be of the same type. The ultrasound will be generated by a single transducer, such as a pulsed laser beam generating ultrasound ablatively. The sample or transducer array is moved to rapidly form a B-scan, and wideband ultrasound will provide good temporal resolution. Conventional systems only use compression waves, but this project will investigate the use of simultaneous shear and compression wave generation and signals mode-converted when scattered using multi-modal emitters and receivers. It will also consider surface modes such as Rayleigh waves. For example, lasers will generate wideband ultrasonic Rayleigh, shear and compression modes simultaneously over predictable and wide angular ranges, and EMATs can be used to detect such varied signals. There are significant differences in propagation characteristics between these wave modes.

The raw data collected is passed through a sequence of processing algorithms with the aim of improving B-scan signal-to-noise ratio (SNR) and creating a cross-sectional (through-thickness) image of the sample. The operator can then clearly see any detected defects, and to some extent, these are automatically identified and presented as a list of scattering positions. Automated sizing is possible if the defect shape is known, but for defects of any but a few standard shapes, there is no single measure of size and hence this cannot be done.

1.3 Time of flight diffraction

The basic principle of ultrasonic testing for defects is for an ultrasonic acoustic wave to be generated at a transducer, scatter off the defects, and then be received at the same or another transducer. TOFD is a mature branch of ultrasonic inspection, and there are also a large number of related techniques. In order to make best use of the existing knowledge, a literature review was conducted, and given in this section is a brief summary of the review.

1.3.1 Amplitude based sizing

Prior to TOFD, most defect sizing techniques were amplitude based. Consider the pulse-echo technique as an example, often used in a single-probe form. This uses a single transducer which sends and receives data from multiple points on a sample surface, often considering signals coming from perpendicular to the surface, although other angles are also common. This technique can build up a map of the sample interior fairly quickly, with defects orientated perpendicular to the probe direction producing strong reflections (strong reflections are also seen due to sound reflection from an edge, known as the corner effect). Defect depth along the probe direction can be determined using time of flight of ultrasonic signals, but sizing is performed using something like the 6dB drop technique [3, 9].

The 6dB drop technique starts with the probe beam over the defect, and moves to the edge until the signal amplitude is reduced by 6dB from its maximum value at the centre, which is assumed to be the edge of the defect. This is only true for a defect with straight edges and away from defect corners, and hence this technique is not particularly accurate. It is generally acceptable for large planar defects operated on by a normally orientated transducer, but not if the defect is of a comparable size or smaller than the beam diameter. The similar 20dB drop technique improves accuracy, but is not a solution to the problem, nor is the distance amplitude correction (DAC) technique, which uses calibration reflectors to relate returned amplitude to defect size. These techniques are affected by the same issues regarding defect orientation, shape, and roughness, as well as variations in coupling [3, 9].

There are more specialist amplitude based techniques, some using double-probe operation [3,9]. The tandem technique uses two 45° (or other appropriate angle) shear wave probes facing the same direction and moved along the plane in which they jointly lie. The rear probe transmits and the front probe receives any sound reflected from flaws within the illuminated area. With relative probe motion and some knowledge of the geometry of the sample, it is again possible to calculate where the flaws lie within the defect [10]. However, the general problem remains that using amplitude as the primary measure of defect size can be very inaccurate.

1.3.2 Time of flight based techniques

TOFD was developed to help in the NDT of nuclear power plants [1,11–13], used to successfully test pressure vessels and nozzles, and turbine and generator components, and is useful for general weld inspection and plant monitoring [1]. TOFD locates (using the intersection of elliptical loci [2]) and sizes defects either within the bulk of a sample or surface breaking cracks [3, 14]. Often, operation is against vertical cracks, but cracks that are titled and/or skewed can also be sized successfully, although the procedure is more complicated [1]. Unlike the other ultrasonic methods used at the time, rather than the reflected (echo) amplitude [15], it relied on the time of flight of a wave after

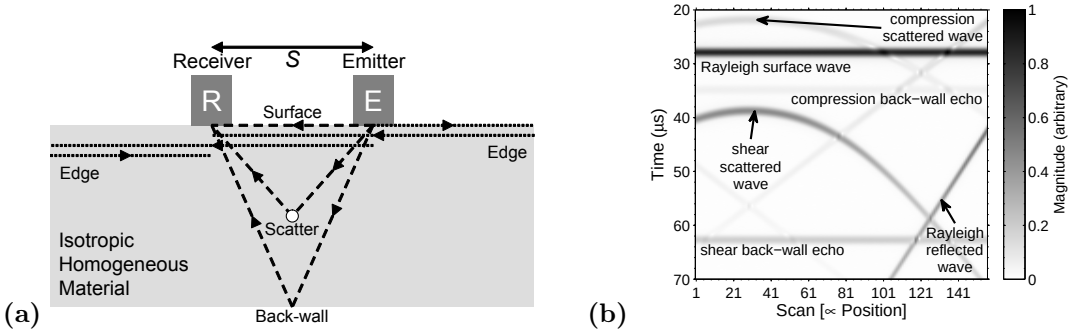


Figure 1.1: The diagram (a) depicts the classic TOFD setup (without mode conversion) representing a single measurement, or A-scan, although usually shown with a crack rather than a point-scatterer. Also shown are surface waves reflected from the edge of the sample. A B-scan usually involves taking multiple A-scans at shifted emitter-receiver positions, retaining the emitter-receiver separation and angle, and using a constant scan motion either parallel or perpendicular to the line between the emitter and receiver. A simulated example is given (b) for a scan motion parallel to the line S, including the Rayleigh wave reflected from the edge. Defects can be detected by a combination of changes to the anticipated signals and the addition of extra signals (diffracted, reflected or otherwise).

diffraction by a defect tip, which is not subject to variations due to coupling factors, reflectivity, angle, or roughness of the defect [6, 11, 15]. TOFD is covered by British Standard BS7706:1993.

TOFD uses a separate emitter and receiver, that are held in a fixed geometry, as they are scanned along a surface of a test sample, figure 1.1a. A pulse of ultrasound is generated at each scan position (an A-Scan), ultrasonic waves propagate into the bulk of the sample, scatter off any defects present, and are subsequently received, before moving to the next scan position. Diffraction causes energy to be spread over a wide angular range; TOFD aims to detect weak diffracted waves arising at edges and tips, which can be used to size defects accurately, whilst avoiding stronger specular reflections which often mask them. Standard procedure is to optimise the probes for a particular defect depth [1]. Waves may also be reflected or shadowed in certain geometries, such that the receiver does not observe expected returns.

If the time-of-flight for each return could be accurately measured, and the material assumed to be homogeneous and isotropic, an ellipse could be plotted (the line along which the sum of the distances to two foci, the emitter and receiver, is constant), on which the ultrasound scatterer would lie. If the defect is assumed to be mid-way between the emitter and receiver, the depth can be estimated (the lowest point on the ellipse). The assumption of a defect at the mid-way point can be reached by lateral displacement of the transducers to achieve the minimum time delay [13]. Early work talks of using multiple probe positions to form multiple ellipses from which to size a defect [14]. Unfortunately, when examining a single A-scan, it is very difficult to determine what a signal might correspond to [16]. The development of the B-scan provided sufficient data to both detect and size defects, and to some extent characterise defects. A B-scan consists of several A-scans, from the TOFD process, stacked together to form an image [1], an example of which is figure 1.1b. Different features exhibit different behaviour as the emitter-receiver pair are moved. In this way, it is possible to identify the features of interest, even with a poor SNR, and hence extract accurate times-of-flight. For convenience here, it is assumed that scan position is along the horizontal axis of the image, and time is along the vertical axis (so that the A-scans are stacked such that each column is a different A-scan, and a row represents the same time across the

A-scans). Thus the data received is the position of the emitter-receiver pair, and the signal at each position, with colour (here luminance on a grey scale) representing signal amplitude or magnitude, in this case magnitude.

If there is an emitter position vector $\mathbf{E}$ and a receiver position vector $\mathbf{R}$, then for the B-scan to make sense, $\mathbf{E} - \mathbf{R}$ is constant even though both positions are functions of time or scan number, $\mathbf{E} = \mathbf{E}(t)$ and $\mathbf{R} = \mathbf{R}(t)$. In general, the motion of the emitter and receiver can be described by:

$$\mathbf{E}(t) = \mathbf{E_0} + t\mathbf{u} \tag{1.1}$$

$$\mathbf{R}(t) = \mathbf{R_0} + t\mathbf{u} \tag{1.2}$$

$\mathbf{E_0}$ and $\mathbf{R_0}$ represent initial positions, $\mathbf{u}$ represents the "velocity" of the scan, and t represents time or scan number (depending on whether the scan is being treated as continuous or discrete). Generally, the scan will move in a direction either parallel or perpendicular to $\mathbf{E} - \mathbf{R}$. B-scan motion is usually along the line containing the emitter and receiver, also known as a transverse or parallel scan, and motion perpendicular to this can be referred to as a D-scan [1,3], also known as a longitudinal, perpendicular, or non-parallel scan. However, B-scan is often used in the literature to refer to both directions [1,10]. The terms B-scan, D-scan, and C-scan are also used within pulse-echo imaging. In that case, a C-scan is a top view image, a B-scan is a transverse section taken in the scanning direction, and a D-scan is a longitudinal view taken orthogonal to the scanning direction. Wherever B-scan is referred to within this study, it is always the TOFD B-scan that is being referred to unless explicitly stated otherwise.

It is common to see a linear feature extending along the position axis; this is a surface wave of some form. There may be several such features for each possible wave mode at different times, but they are easy to spot. Each feature is constant unless there is a surface breaking defect, since $t_c = |\mathbf{E}(t) - \mathbf{R}(t)|/c$ (where c is the wave speed) is constant with respect to time.

As the scan moves down the sample, curved features (often referred to as characteristic arcs [16], parabolas to 2nd order [17, 18]) are seen when the emitted wave scatters off a defect at position vector $\mathbf{D}$. Assuming an isotropic and homogeneous medium, the distance the wave travels, d, and the associated time, t_d, are:

$$d = |\mathbf{D} - \mathbf{E}(t)| + |\mathbf{D} - \mathbf{R}(t)| \tag{1.3}$$

$$t_d = \left(|\mathbf{D} - \mathbf{E}(t)| + |\mathbf{D} - \mathbf{R}(t)|\right)/c \tag{1.4}$$

These curves may be repeated at different times corresponding to different wave modes. If the defect is completely within the bulk (not surface breaking), there will be diffraction from both the top and bottom, giving a measure of the depth extent [6].

In addition to the standard direct, back-wall reflection, and diffracted (from a defect) waves incident on the receiver during an A-scan, there is also a wave (depicted in figure 1.1a) going from the emitter, to the sample edge (along the surface), and back to the receiver (example geometry is depicted in figure 6.2). This forms angled (but still straight) lines when performing a B-scan, and can cause significant interference to the parabola observed due to the scatterer, the feature (defect) of interest. Such lines can be seen in the example B-scan of figure 1.1b.

The original version of TOFD only concerned itself with the fastest bulk waves, which are

longitudinal (compression) waves, as this avoided ambiguities in mode identification [1, 11]. Other wave types were eliminated through a combination of emitter-defect-receiver geometry choices, time gating (since compression waves are approximately twice as fast as shear and Rayleigh wave modes [19], as is the case for steel and aluminium), and, significantly, transducer choice. The piezoelectric transducers that were used in the original work would have primarily emitted and received only compression waves, and other wave types would be heavily attenuated due to a fluid coupling layer (which cannot support shear wave modes) between the sample surface and the transducer .

The factors which make TOFD stable against lateral probe movement, particularly the small change in depth of the ellipse at its lowest point, also make any estimate of the lateral position much worse than any depth estimate [13]. Assuming no change of pulse shape, it is possible to estimate a given point on the steeper parts of a waveform to 0.01 of a wavelength. The accuracy in estimating depths is worse due to geometrical factors, 0.03-0.05 of a wavelength in ideal circumstances. However, changes in pulse shape due to short-term variations in the crack profile (dispersion and differential absorption are not important in steel below 10MHz), cause the returning pulse to be the sum of varying components of slightly different phase, despite which accuracies are usually still better than 0.3 wavelengths [13].

1.3.3 Developments in time of flight diffraction

Since its first development, there has been considerable work into analysing the data generated by a TOFD B-scan, and into making that data easier to interpret. The data is considered difficult to interpret due to the mode-conversions, geometric reflections, and other noise signals [14], despite the characteristic arc shape due to a defect in a B-scan, which makes crack determination far easier than for an A-scan [16]. The use of mode-converted signals for sizing was considered, but not used initially due to a desire for simplicity [2].

1.3.3.1 Improvements to sizing

Early work pioneered the use of multiple probe setups (and orthogonal scans) for TOFD [20], and indeed, early tests of TOFD used an array of probes [1]. Since probes tend to have a favoured direction, this gives sensitivity over a wide range of depths and reduces the need for raster scanning. Locating defects in three dimensions was also introduced by finding the common intersection of elliptical loci constructed from all observations of the feature. However, this was not imaging since it used identified signal returns, and hence if there was ambiguity in which signal returns were those of interest, there would also be ambiguity in the position of the defect. It is not entirely clear how elliptical surfaces (ellipsoids) which were not coincident at a single point were dealt with, but it appears to have been by looking at discrete regions with the highest density of crossing points. This does resemble the modern imaging method in many ways, but in terms of display, it was taken no further than the B-scan; cross-sectional images do not seem to have been used, possibly due to the computational constraints of the time.

Highly automated systems (although also highly specialised to particular sample geometries) incorporating pulse-echo and TOFD collections with synthetic aperture focusing techniques (SAFT) to improve lateral resolution were developed [21–23]. Pulse-echo seems to have first been used to identify defects, and then "virtual boxes" placed around those of interest for TOFD sizing. Mode-converted signals were acknowledged as causing defect misinterpretation, but able to provide

information on defect character (the extent to which this character was explored is not clear).

Other early work superimposed ultrasonic data onto an outline of the sample, such as a pressurised water reactor (PWR) inlet nozzle, based on probe position, beam direction, and sample geometry [14]. Unfortunately, the output could appear unclear, with some ultrasound indications outside the outline, as well as other issues. SNR was low due to the many sources of noise such as geometric reflections and grain noise, and the defect position was certainly not clear; a skilled human operator was still required. The superimposition could be combined with software allowing an operator to draw on ellipses, and from this, the defect position/size could be calculated. As with prior attempts, this could be a very difficult task if there were multiple reflections or mode-conversions.

Mathematical modelling, including sample geometry, probe movement, and beam direction, allowed for optimisation of inspection techniques and formed the basis for data analysis routines to size and characterise defects [22,23]. The range and amplitude of significant peaks are extracted algorithmically. Like similar scanning processes [10, 20], at least two orthogonal scans are used. However, it appears that again, cross-sectional images were not generated (possibly due to the computational constraints of the time).

Classic TOFD is also being developed for use in more complex geometries, such as a steam turbine rotor [24].

1.3.3.2 Defect characterisation

Other studies examined the critical size a defect needed to be to reject a sample, and considered the errors in TOFD when sizing [25]. Error in defect depth was found to increase with defect depth, meaning that the best resolution could be obtained for shallow defects, although in practice, defects too close to the transducer could overlap with the emission or surface wave [10, 26]. Again curve-fitting and SAFT were considered to find location and depth; other workers have noted that it is necessary to make a correction to measured time delays to allow for the difference in waveform between the lateral and diffracted signals and for the change in form of the diffracted signal with depth [26]. Surface breaking cracks can be identified by having a missing signal from the top part of the crack (nearest the signal), but this overlaps with cracks for which the signal from the top part of the crack is poorly resolved. Beam width results in a defect signal appearing before and after the joint centre of the transducers has reached the defect, artificially giving the defect additional width. Other workers have noted that very oblique angles make accurate defect positioning difficult [16], leading to large errors in defect position for small errors in timing and probe position. Other errors might be due to noise leading to a low probability of detection for the defect [27]. In addition, precision in measuring defect length is related to beam width, but precision in measuring defect depth is related to the wavelength of the ultrasound [28]. Practical considerations include the errors in measurement of signal arrival times and transducer separation, variations in ultrasound speed, and variations in the thickness of the material (if the back-wall echo is being used) [29].

The characteristic arc shape of a parabola in the B-scan, caused by a point-like scatterer, is mentioned many times in the literature [14, 26, 29–32] (although it is also noted that if this shape overlaps with the lateral wave, complicated interference effects often appear [26]), culminating in the use of a Hough transform to search for that shape within the scan [17, 18, 33, 34]. The length of the arc for a defect parallel to the surface (but within the bulk) is equal to the sum of the length of the flaw and the length of the arc from a point flaw [28]. This would suggest that there is benefit

in highlighting the shape through a matched filter operating on the B-scan image and displaying the B-scan to the operator. Additionally discussed is the effect that the defect has on the usual signal returns, namely the lateral wave and back-wall echo [29]. The automated sizing of defects has also been performed for D-scans, by more general curve-fitting techniques [35].

Once sizing had become more accurate, attention was turned to other facets of defect characterisation, in particular, determining if a signal was the top or bottom of a crack [36] and discriminating between different defect types [37] (it seems that this had only been done previously to the extent that an appropriate sizing technique could be chosen [21]). For the former, the phase of the signal was extracted using phase based cross-correlation with a matched filter, the constant phase offset being of sole interest since the linear change would be due to differences in the time of arrival. Line fitting was used to get the phase linear relationship. When performing TOFD, the phase of a signal will be lagging by $\pi/4$ if scattered from the bottom of a crack, and leading by $3\pi/4$ if scattered by the top, relative to a signal that has not been scattered [1, 3, 36]. In work not involving TOFD, defect imaging, feature extraction, and pattern recognition methods to distinguish significant defects such as cracks from volumetric flaws such as slag inclusions and porosity, were in use [3]. Although TOFD was not used in that work, it is still of interest as the techniques could be applied to TOFD. In this case, pulse-echo and tandem methods were employed in a raster scan using multiple angles of illumination to give a specular reflection and backscattering information. Signal returns can be associated with positions in the sample and features used to discriminate between defect types. The types of defects that could be distinguished were single crack-like (rough or smooth) surfaces, single volumetric, composite flaws, and multiple flaws (small and rough multiple cracks are indistinguishable from multiple volumetric flaws).

1.3.3.3 Synthetic aperture focusing technique

As computational capabilities improved, the usage of advanced signal processing became more prominent [32, 38]. SAFT has been used in NDT in several studies [20, 22, 23, 39, 40]. SAFT adds non-linear time shifts to the radio frequency (RF) data and uses coherent summation of elliptic functions in the spatial domain, the parameters for which are functions of the emitter and receiver positions and the time-of-flight of the scattered signals [22, 38].

Pulse-echo data is relatively simple to convert into location and depth data based on probe position and signal return time of flight respectively, and this can be done in conjunction with SAFT and then put into a cross-sectional image, overlaying a diagram of the sample. Flaw-profiles were highlighted by using curve-fitting techniques to outline the shape of the echo. Also mentioned was the usefulness of reducing the data to a list of defects with estimates of position and size [32], although it is not clear that this was in any way automated at such an early stage.

Some versions of SAFT use Gaussian aperture weighting (apodisation) to minimise spurious side-lobes [39], and Fourier transform (FT) versions also exist (FT-SAFT, although other names have been used) [38, 41–43], some considering the use of mode-converted waves although only a single mode or mode-conversion at any given time [42]. Work was done using a sparse linear array and a technique equivalent to what is now known as full matrix capture (FMC) to create an image in the spatial domain using SAFT [38], differing from most of the other SAFT work that used an array of transducers, but only ever received on the transducer that was the current emitter. Work was also done to incorporate the spatial impulse response (SIR) of the transducer into SAFT-like processing to improve resolution [44], and to measure the effect of spatial sampling

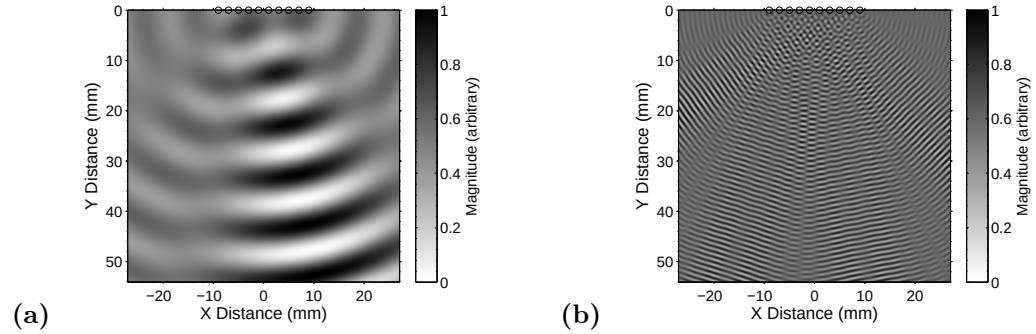


Figure 1.2: This is a simulation of a 10 element array of transducers (circles at the top of the images), each element separated by 2mm, with delays to produce an ultrasonic beam at an angle of 11° propagating at 6000m/s. First consider a monochromatic emission at 0.6MHz (a); there is only a single beam, and it is this direction that will produce strong signals, as the element separation is a fifth of the wavelength of 10mm. Now consider a monochromatic emission at 6MHz (b); there are multiple lobes from which a strong signal will be received, as in this case, the element separation is double the wavelength of 1mm and spatial aliasing has occurred. Temporally sharp broadband pulses do not produce this effect.

on the signal-to-clutter ratio of images [45]. Spatial undersampling creates spatial aliasing, which effectively means that there are several lobes of high sensitivity rather than just one main lobe. Consequently, strong signals may appear to come from somewhere they are not expected, as can be seen in figure 1.2.

The output of SAFT is a function of the response of the scatterer, and only the contour of the scatterer that is visible from the emission and reception apertures is imaged, as is intuitively expected. The oscillations of the ultrasonic pulse can be seen if only the real part of the signal is used. Using the full analytic signal, including real and imaginary parts, produces sharper images by including the phase information; then the magnitude of the output image is taken to remove the oscillations of the signal. Taking the magnitude of the analytic signal prior to imaging is faster, as the SAFT algorithm receives only the envelope of the actual signal, and hence half the amount of data. As with taking the magnitude of the image, the oscillation of the signal is not present, but the resolution is inferior to using the complex data as the input for the imaging algorithm, and hence is not advised unless the phase is unclear or corrupted [42].

1.3.3.4 Non-standard probe configurations

It was established that defect detection could be performed using an array of standard normal incidence piezoelectric probes operating on defects away from the normal [46, 47], using the same single emitter and multiple receivers setup as in this project. This was referred to as normal probe diffraction (NPD), and implicitly used only compression waves since time-gating removed everything after the first back-wall echo, and further echoes could not be distinguished due to multiple scattering events. The amplitude of the diffracted signal was small relative to a back-wall echo, and required signal processing (band-pass and matched filtering) to make it more visible, as could be expected when operating outside of the beam of the transducer. Multiple ellipses, one for each transmitter-receiver pair, with a transmitter as one foci and a receiver as another, were used to locate the defect, and thresholding was used to detect the defect peak.

The use of compression waves mode-converted to shear waves (upon diffraction) for TOFD has

been previously established [48]. The use of both compression and shear waves means that if one of the wave modes scattered from a defect is obscured by a direct surface wave, then the other wave mode scattered is likely to be away from the surface wave, and under certain conditions, shear waves can provide greater resolution than compression waves.

Laser-EMAT systems are well established within NDT [49]. For example, such a system has been utilised for thickness monitoring using shear and LS mode-converted waves [50], for ultrasonic weld inspection [51], for detection of surface-breaking cracks [52], and for directivity measurements of a laser line source [53]. Such a system has also been used for inspection of hot, moving steel [4,5]. Some work reports operation on steel at up to 1100°C, with laser generation of ultrasound, and water-cooled EMATs for reception which are momentarily placed on or near the sample, but not left in contact for longer than necessary to take the measurements [54]. EMAT-laser systems, with the EMAT used for generation and the laser for reception, have been used for many applications. An example is investigating the interaction of Rayleigh waves with surface-breaking cracks for depth gauging, and how similar to Rayleigh waves the guided waves on a curved rail head are in the near-field [55], as well as measuring the waves generated by an EMAT [56]. Laser-laser systems have been used for investigating the interaction of Rayleigh waves with surface-breaking cracks for depth gauging [57, 58], and for estimating the thickness of thin metal sheets [59]. EMAT-EMAT systems are well established, and have been used for thickness gauging [60], crack detection and depth gauging by Rayleigh waves [61–64], for the integrity inspection of steel [65], and numerous other purposes. All of these sorts of systems have the potential to simultaneously monitor for internal and surface defects if designed to do so, although it is common for either bulk or surface defects to be the focus of a technique.

1.3.3.5 Imaging

Modern techniques can trace their origins back to TOFD. The total focusing method (TFM) uses a dense array of transducers, and captures data for every possible transmit-receive pair in a process known as full matrix capture (FMC) rather than the more conventional phased array. All the data is used although typically only for a single wave mode. Every point in the image is focused on via post-processing the FMC data; effectively, it is simply taking the received signal data at the time corresponding to a signal scattered from that point, for each emitter-receiver pair [66]. Although of course the hardware for parallel reception is very expensive, the resolution of the image produced is excellent. This process is similar to earlier NDT work [38], which instead referred to acquisition of a complete set, collecting data at each receiver point for every possible transmission point, and processing the data by taking the coherent summation in the space domain of elliptic functions defined by the emitter and receiver positions and the time of flight, for every emitter-receiver pair. There are variations on TFM such as the vector total focusing method (VTFM), which attempts to extract information on defect orientation. Some work on SAFT used complex signal data when forming an image and the magnitude of the image was taken [42], as was later used in TFM [66].

There are also developments incorporating multiple reflections of the signal into the TOFD processing, allowing probe separations of around a metre and wall loss measurement accuracy of ±1mm, for inspection of inaccessible components [67]. Angled shear wave probes are used to avoid mode conversion losses at reflections, and although it isn't stated, the implication is that these are horizontally polarised shear waves. In general, shear waves are known to undergo mode conversion at interfaces, but not horizontally polarised shear waves (assuming they are horizontally polarised

relative to the encountered interface). The technique appears to successfully use specular reflections that have reflected off the back-wall, one, two (with a reflection off the front-wall), or three times (with two reflections off the front-wall), or even more. The arrival times of the multiple reflections can be used in conjunction with a simple equation for the wall thickness and least squares fitting if necessary. Pitting type wall loss can be detected, as it changes the flight time of one or more of the reflections. Iterative numerical techniques are required if performing circumferential beam inspection.

Techniques incorporating FMC and using multiple reflections are under development [68]. The shape of an object is reconstructed using the standard scattered waves from the directly illuminated part of the defect, and the scattered waves from the indirectly illuminated part. The waves first reflect from a planar interface (an ideal back-wall), then scatter from the back of the defect, before reflecting again off the back-wall, and finally being received. This means that the limited view of the transducer array is circumvented; the defect does not have to be scanned from every direction to reconstruct the shape. Unfortunately, the technique as currently presented for NDT work appears to be limited to simulated data (there are similar and much older techniques in other fields, such as geophysics, which have been proven experimentally, but they are of limited interest here), but the factorisation method used for the reconstruction provides impressive results relative to the beamforming method used as a comparison. Only compression waves are considered at the moment, not shear waves, and it will be interesting to see how the technique deals with a signal of interest cluttered by various mode-converted waves.

1.4 Use in development of TOFDI

This chapter has put TOFDI into context, describing existing techniques and other relevant research that is already available in the literature. From this, the importance of the development of TOFDI can be seen, as a technique that can operate in a harsh environment, dealing with a rough and moving sample, that is significantly larger than the samples many other techniques were developed for. TOFDI may have to use transducers that produce multiple wave modes and mode conversions. Multiple wave modes can aid the imaging process, by providing scattered signals that are not obscured by a direct or reflected surface wave or by a back-wall echo. However, they can also hinder the imaging process, as unambiguous identification of which wave mode a signal corresponds to is not always possible, and they will result in more direct and reflected surface waves and back-wall echoes.

Chapter 2

Theory

This chapter describes the theoretical aspects of ultrasound generation, reception, propagation, interaction with interfaces, and scattering from defects. Defects expected, in the sort of applications TOFDI will be applied to, are discussed, including their basic form, and how they occur during production and service of a component. Electromagnetic acoustic transducers (EMATs) and pulsed lasers are presented as useful ultrasound receivers and emitters when dealing with rough (and also hot) samples, including the basics of how they detect and generate ultrasound respectively (as used within this work). Although many of these theoretical aspects are not mentioned explicitly after this chapter, they are included here to put the wave behaviour observed in this work, and the transducer choices made for this TOFDI implementation, within a suitable context.

2.1 Ultrasound in solids

The work within this study assumes a homogeneous, isotropic medium. Inhomogeneous and/or anisotropic material is not considered, although material such as textured austenitic weld material or cladding is of interest to NDT inspection [30, 69, 70]. It is also assumed that the stresses due to wave propagation are well within the elastic regime of the material; without this constraint, the propagation characteristics would vary even within a homogeneous isotropic medium [3].

Index notation with the summation convention is used in this section. Additionally, $a_{ij,j}$ is $\partial a_{ij}/\partial x_j$, and $a_{j,ji}$ is $\partial a_j/\partial x_j \partial x_i$, with the derivative always being spatial. Time derivatives are given in long form, d/dt, or da/dt as $\dot{a}$ if appropriate.

2.1.1 Waves in elastic solids

The equations representing wave propagation in infinite elastic solids require some assumptions that are generally true for all solids. These include the conservation of mass, momentum, angular momentum, and energy [71]. In addition, a non-linear set of equations is being forced to linearity by assuming, for example, small deformations. The equations will become non-linear if approximations made are broken, but this is not of any concern to this work [3].

In the general anisotropic elastic case there are from 3 to 21 elastic constants depending on the degree of anisotropy. If homogeneity and isotropy are assumed, the elastic constants can be reduced to the 2 Lamé constants, λ and μ. The shear modulus, μ, sometimes also written as G [72], is the ratio of sheer stress to shear strain. Elastic moduli vary with temperature and material properties

such as hardness, grain size, applied load, and internal stresses [3].

Assume a mass density per unit volume, ρ, and a body force per unit mass of material, f_i. In terms of displacements, u_i, the kinematics of an arbitrary point is governed by the local strain-gradient field, $u_{i,j}$ (solutions can be shown to be unique if boundary conditions are specified) [71,73]:

$$(\lambda + \mu)u_{j,ji} + \mu u_{i,jj} + \rho f_i = \rho \ddot{u}_i \tag{2.1}$$

For simple tension and pure shear loadings, additional elastic constants can be defined. Young's modulus, E, is the uniaxial stress over the uniaxial strain. The bulk modulus, K, is the pressure increase for a given relative decrease in volume. Poisson's ratio, ν, is the lateral contraction per unit breadth divided by the longitudinal extension per unit length under an applied longitudinal stress [72]. The longitudinal modulus, M, is the ratio of axial stress to axial strain in a uniaxial strain state [72]. Relating these constants to the Lamé constants [1, 71]:

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu} \tag{2.2}$$

$$K = \lambda + \frac{2}{3}\mu \tag{2.3}$$

$$\nu = \frac{\lambda}{2(\lambda + \mu)} \tag{2.4}$$

$$M = \lambda + 2\mu \tag{2.5}$$

A general plane wave, with amplitude, A_i, wave normal, n_i, and phase, $\psi = n_k x_k - ct$, is given by [71]:

$$u_i = A_i f(n_k x_k - ct) \tag{2.6}$$

A solution is the simple harmonic plane wave, of magnitude and polarisation $\mathbf{A_0}$, angular frequency, ω, wave-vector, $\mathbf{k}$, phase, Φ, at a position, $\mathbf{r}$, from the emission point, $\mathbf{r_0}$:

$$\mathbf{A} = \mathbf{A_0} e^{j(\omega t - \mathbf{k}\cdot(\mathbf{r}-\mathbf{r_0}) + \Phi)}$$

Usually the two wave sources considered are the plane wave, from a source at infinity, and the spherical wave, from a point source. Plane waves do not spread geometrically, whereas spherical waves experience the standard inverse square loss of intensity. Real sources can be more complicated.

2.1.2 Bulk waves

The general condition of normal and shear loads applied to a surface of a half-space results in both compression and shear waves being generated [71]. These are the two possible wave modes in the bulk of a homogeneous isotropic solid [1], but in homogeneous anisotropic solids, there are three distinct modes; one quasi-compressional and two quasi-shear [74]. The wave modes for an anisotropic medium will be neglected as a homogeneous isotropic medium is used throughout this work.

A compression wave (irrotational volumetric disturbances for which the displacements are parallel to the propagation direction) will propagate at c_c [71]:

$$c_c = \sqrt{M/\rho} \tag{2.7}$$

Shear waves (equi-voluminal rotational disturbances for which the displacements are transverse to the propagation direction) propagate at c_s [71]:

$$c_s = \sqrt{\mu/\rho} \tag{2.8}$$

Indeed, the general relationship between the speed of sound, c, in a solid, its density, and the associated elastic constant, χ, is [3]:

$$\chi = \rho c^2 \tag{2.9}$$

Since $0 \leq \nu \leq 1/2$, $c_c > c_s$, and $0.55 > c_s/c_c > 0.49$ [75], from the ratio of the wave speeds [71, 76]:

$$k = c_c/c_s = \sqrt{(\lambda + 2\mu)/\mu} = \sqrt{(2 - 2\nu)/(1 - 2\nu)} \tag{2.10}$$

It is clear that the polarisation for a shear wave can lie in more than one direction, and shear waves are commonly split into shear horizontal (SH) and shear vertical (SV). SH waves are defined as having a displacement parallel to the surface of the medium in which they are propagating, whereas SV waves have a displacement perpendicular to the surface. The shear waves can have a displacement at any angle that is perpendicular to the direction of propagation, but choosing two convenient orthogonal axes means that any transverse wave can be constructed by combining SH and SV waves. The separation of polarisations becomes important at boundaries, at which the wave behaviour (and how it can be detected) varies depending on the polarisation [1].

2.1.3 Waves at interfaces

At any interface (any abrupt change in acoustic impedance such as a boundary or defect), the following can occur [6, 71, 74]:

- Transmission through the boundary where refraction will occur;
- Reflection from the boundary with a change of amplitude and phase;
- Conversion of elastic energy at the boundary into a different bulk wave or a surface wave;
- Attenuation due to inhomogeneity at the boundary.

A smooth boundary causes reflection, whereas a rough boundary causes scattering [75]. If the boundary is a free surface, pure reflection will occur [71]. If the second medium cannot support the appropriate particle motion, such as at a solid/fluid boundary, the wave may totally reflect or may be partially mode-converted to a compression wave [74].

For plane progressive non-attenuating waves, if the material density is ρ and the wave speed in that material is c, the acoustic impedance, z, is given by [3, 75]:

$$z = \rho c \tag{2.11}$$

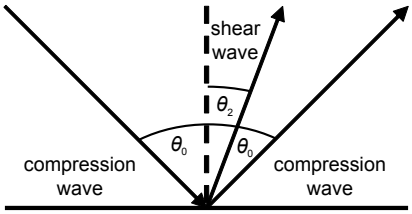


Figure 2.1: The reflected ($\theta_1 = \theta_0$) and mode-converted wave (θ_2) angles for an incident compression wave (θ_0).

The effect of an attenuating medium (attenuation factor α at angular frequency ω) is to make the impedance complex [3]:

$$z = \rho c(1 - j\alpha/\omega c) \tag{2.12}$$

For a plane wave in a material of acoustic impedance, z_1, incident on a perpendicular smooth boundary with a material of acoustic impedance z_2, the fraction of the amplitude reflected, R, and transmitted, D, is [74, 75]:

$$R = \frac{z_2 - z_1}{z_2 + z_1} \tag{2.13}$$

$$D = \frac{2z_2}{z_2 + z_1} \tag{2.14}$$

Phase reversal ($R < 0$) always occurs if reflected from a sonically softer material, defined as a material with a lower acoustic impedance [75]. Air is so sonically soft that the metal-air boundary can effectively be considered a free boundary [3].

Characteristic of wave-boundary interactions for elastic waves in solids, is the occurrence of mode conversion. Although different wave modes are uncoupled in the bulk of the material, coupling occurs at boundaries through the boundary conditions, such that an incident wave is converted into two waves on reflection [71]. A plane compression wave, amplitude A_0, incident to a free boundary at angle θ_0, will result in a reflected compression wave (amplitude A_1, angle θ_1, speed c_c) and a mode-converted shear wave (amplitude A_2, angle $\theta_2 < \theta_1$, speed c_s), with polarisation perpendicular to the boundary (SV) [6]. For the wave not converted, the reflection is specular (angle of incidence equals the angle of reflection, $\theta_1 = \theta_0$), and the mode-converted wave follows Snell's law ($\sin\theta_2/\sin\theta_0 = c/c_0$) [6], which can be confirmed using the Huygens-Fresnel principle. These angles are depicted in figure 2.1 for $k = c_c/c_s = 2$. There are critical angles past which reflected waves can disappear [71].

The amplitudes are more complicated [6, 71, 75, 77, 78]:

$$A_1 = A_0 \frac{\sin(2\theta_0)\sin(2\theta_2) - k^2\cos^2(2\theta_2)}{\sin(2\theta_0)\sin(2\theta_2) + k^2\cos^2(2\theta_2)} \tag{2.15}$$

$$A_2 = A_0 \frac{2k\sin(2\theta_0)\cos(2\theta_2)}{\sin(2\theta_0)\sin(2\theta_2) + k^2\cos^2(2\theta_2)} \tag{2.16}$$

These formulae are plotted in figure 2.2a. It is also worth noting that in this case $k = c_c/c_s$ and this can be restated in terms of Poisson's ratio ν such that $k = 2(1-\nu)/(1-2\nu)$. It is then clear that the reflected and mode-converted ratios are only a function of Poisson's ratio. Shear

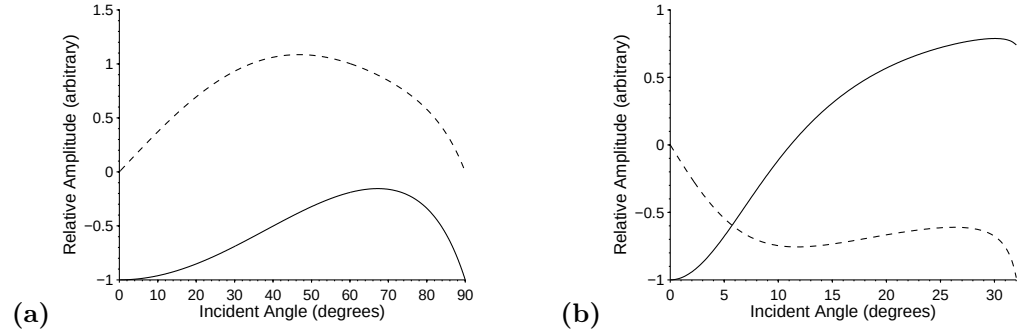


Figure 2.2: The relative amplitude of a reflected (solid line) and mode-converted (dashed line) wave for an incident compression wave (a) and an incident shear wave (b).

wave amplitude can be greater than unity, but this does not violate conservation of energy, since energy transfer is a function of wave speed as well as amplitude, and shear waves are slower than compression waves [71].

For an incident shear (SV) wave (amplitude B_0), a reflected shear wave (amplitude B_1, speed c_s), and a mode-converted compression wave (amplitude B_2, speed c_c), the amplitudes are [6,71,78]:

$$B_1 = B_0 \frac{\sin(2\theta_0)\sin(2\theta_2) - k^2\cos^2(2\theta_0)}{\sin(2\theta_0)\sin(2\theta_2) + k^2\cos^2(2\theta_0)} \tag{2.17}$$

$$B_2 = B_0 \frac{-k\sin(4\theta_0)}{\sin(2\theta_0)\sin(2\theta_2) + k^2\cos^2(2\theta_0)} \tag{2.18}$$

These formulae are plotted in figure 2.2b. The phase shift observed (inversion of the amplitude) is necessary for the direction of displacement to be continuous at the boundary. At 45°, an SV wave does not mode convert and is reflected as a pure SV wave. After the critical angle for an incident shear wave, the mode-converted compression wave becomes a surface-skimming compression wave [6].

A reflected compression wave has a constant phase shift for all angles, whereas a reflected shear wave has a π phase shift away from the critical angle, but a variable shift near it [6]. In addition to the free boundary equations given here, there are also equations available for an interface between two liquids, the interface between liquids and solids, and the interface between two solids [75].

A shear wave with polarisation parallel to the boundary (SH) is reflected without mode conversion, and the amplitude does not change (there is the same π phase shift however) [6,71,75,78]. When compression and SV waves are incident on a free boundary with very specific ratios, they can reflect as themselves [71]. However, this behaviour is not of particular interest and will not be considered further here, being more of use in the study of plates.

Diffraction occurs when a propagating wave encounters an interface that is not a reflecting plane, and is most pronounced if the wavelength is the same order of magnitude as the diffracting object. The Huygens-Fresnel principle states that each point on an advancing wavefront is the source of a new wavefront, and the summation of these wavelets gives the amplitude at any given point. Diffraction can be calculated via this principle; simple cases can be solved analytically using Fraunhofer diffraction for the far field and Fresnel diffraction for the near field, but more complicated cases must be solved numerically. Fraunhofer diffraction shows that the far-field diffraction pattern is the spatial Fourier transform of the aperture, as a consequence of the parallel-

rays approximation, and moving an aperture changes only the size of the diffraction pattern. Fresnel diffraction incorporates the curvature of the wavefront, and moving an aperture changes the diffraction pattern in size and shape. Diffracted signals have less amplitude than reflected signals as the energy is spread out in all directions.

2.1.4 Surface acoustic waves

Surface acoustic waves (SAWs) must be generated to satisfy the stress-free boundary conditions [1]. SAWs have an amplitude, A, that decays exponentially with depth, d, hence the energy, E, is confined to the surface. Generally the amplitude is of interest rather than the energy, and $A^2 \propto E$. For an EMAT, the velocity is measured, and $v \propto A$. For a wave with surface amplitude, A_0 ($d = 0$), and wavelength, λ, the amplitude is [1, 6, 71]:

$$A = A_0 e^{-d/2\lambda} \tag{2.19}$$

If the boundary conditions are for an elastic half-space constrained by a rigid, lubricated boundary, surface waves cannot exist [71, 74, 79]. If there is a liquid on the surface, such as a coupling medium if contact piezoelectric transducers are being used, wave energy is continually lost to the liquid [75]. A plane free boundary of a homogeneous isotropic solid can support SAWs, including the lateral (creeping) wave behaving as a surface skimming compressional wave, Love waves (surface skimming shear waves with a displacement into the plane), Rayleigh waves, and Lamb waves [74]. Some of these waves only propagate in specific geometries; Lamb waves propagate in solid plates, whereas Rayleigh waves require a surface that forms part of a semi-infinite medium [74, 79]. There is a dispersion curve that describes the transition from Lamb to Rayleigh waves as the frequency of the wave and the thickness of the medium changes, hence they are often described as Rayleigh-Lamb waves.

This work considers relatively high frequency waves in samples of significant thickness, and is therefore very much in the Rayleigh regime, so much so, that over the frequencies of interest, the Rayleigh wave can be considered to have no dispersion at all, as the variation of phase speed and group speed with frequency (for a given large thickness) is very small [74].

EMATs can be used to generate Rayleigh and Lamb waves [80]. Rayleigh waves can also be generated when waves are reflected at a boundary, at the third critical angle. At the first critical angle, θ_1, no compression waves are transmitted (speed c_{c2}), only reflected (speed c_{c1}), leading to an increase in reflected compression intensity, but shear waves are both transmitted (speed c_{s2}) and reflected (speed c_{s1}). At the second critical angle, θ_2, no shear waves are transmitted either, only reflected, leading to an increase in reflected shear intensity. At the third critical angle, θ_3, Rayleigh waves (speed c_R) propagate along the boundary, leading to a decrease in reflected wave intensity [3]:

$$\sin\theta_1 = c_{c1}/c_{c2} \tag{2.20}$$

$$\sin\theta_2 = c_{c2}/c_{s2} \tag{2.21}$$

$$\sin\theta_3 = c_{c1}/c_R \tag{2.22}$$

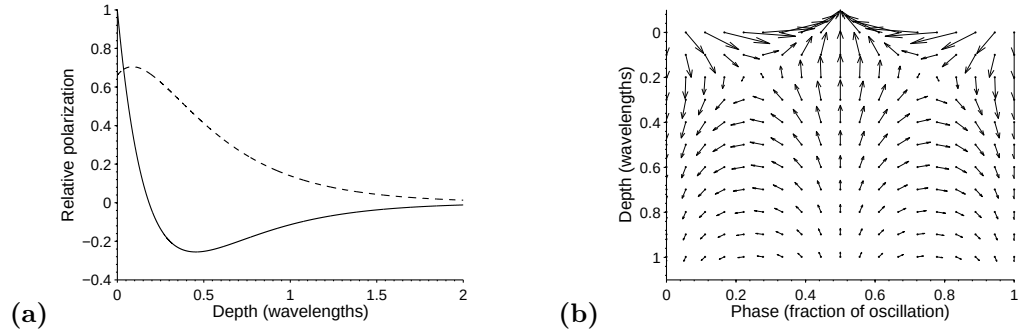


Figure 2.3: The Rayleigh wave polarisations for a 1MHz wave, with wave speeds $c_R = 2987$m/s, $c_c = 5932$m/s, and $c_s = 3226$m/s. The polarisation amplitudes (a) are relative to each other, and given as $\xi(z)$ (solid line) and $-\eta(z)$ (dashed line). The polarisation direction at a range of depths and fractions through the oscillation (b) are also relative to each other, and do not represent absolute displacement amplitudes.

Rayleigh waves have an elliptical particle oscillation, retrograde with respect to the direction of propagation (counter-clockwise for a wave travelling to the right), with a normal displacement approximately 1.5 times larger than the tangential displacement [6,71,74,79]. The particle movement loses amplitude and becomes more circular at increasing depth [75]. At a certain depth (approximately 0.2 wavelengths) the particle rotation direction reverses [71]. Clearly there is a transition region, and indeed there are depths at which the horizontal motion vanishes, although the vertical motion never does [79]. The Rayleigh wave polarisation (figure 2.3), split into components parallel to the propagation direction, ξ, and normal to the surface, η, with zero component parallel to the surface and normal to the propagation direction, are described by the following equations, noting that the values are only relative to each other and not actual displacements [81]:

$$\xi(z) = \frac{k/q_s}{k^2 - q_s^2}\left[\left(k^2 + q_s^2\right)e^{-q_c z} - \left(2q_c q_s\right)e^{-q_s z}\right] \tag{2.23}$$

$$\eta(z) = \frac{1}{k^2 - q_s^2}\left[\left(k^2 + q_s^2\right)e^{-q_c z} - \left(2k^2\right)e^{-q_s z}\right] \tag{2.24}$$

The polarisations are functions of depth, z, and are dependent on the wavenumber, $k = 2\pi/\lambda$, where λ is the wavelength. For Rayleigh, c_R, compression, c_c, and shear, c_s, wave speeds, the decay constants, q_c and q_s are given by [81]:

$$q_c^2 = k^2(1 - c_R^2/c_c^2) \tag{2.25}$$

$$q_s^2 = k^2(1 - c_R^2/c_s^2) \tag{2.26}$$

Rayleigh waves are significantly attenuated by rough surfaces [81–86]. Scattering due to roughness can be from Rayleigh waves to Rayleigh waves of a different propagation direction, or to bulk waves (to a lesser extent), and the scattering is proportional to f^5 for surface roughness that is on a small scale relative to the wavelength, but is less for larger scale roughness or defects below the surface [81–83]. This will force the frequencies to lower values via the attenuation [4]. Rayleigh waves are also slowed by surface roughness [85]. On a similar theme, there is work exploring the Rayleigh wave attenuation and dispersion due to distributions of one-dimensional surface breaking

cracks [86], and speed changes due to surface plane stress [87].

An equation relating the speed of compression bulk waves, c_c, and shear bulk waves, c_s, to the Rayleigh wave speed, c_R, is [1, 6, 71]:

$$\left(c_R^2/c_s^2\right)^3 - 8\left(c_R^2/c_s^2\right)^2 + \left(24 - 16c_s^2/c_c^2\right)\left(c_R^2/c_s^2\right) - 16\left(1 - c_s^2/c_c^2\right) = 0 \tag{2.27}$$

This is simple to solve numerically, and indeed to rearrange such that for any given two wave speeds the third can be found.

Component curvature changes the speed of waves creeping around them. The solutions are not trivial, but for large radius curves, the wave speeds are approximately that of the compression and shear waves, and a surface wave [1].

It is also worth mentioning Head waves, which are generated by continuous mode conversion when a wave passes along a free surface at a speed greater than that of the bulk wave speed in the medium, such as when a surface skimming compression wave generates shear waves as it passes along a surface or down a crack [74]. The Head (shear) wave appears as a straight wave front emanating from the compression wave skimming along the surface. Head waves are very occasionally incorporated into NDT work [65].

2.1.5 Geometric attenuation

Wave energy spreads out as a wave propagates. In many cases it is useful to use the approximation of a uniform omni-directional point-like source. The standard derivation of the inverse square law uses the surface of a sphere. An area element, dS, on a sphere of constant radius, R, is (using spherical coordinates, ϕ is elevation, and θ is azimuth):

$$dS = R^2 sin\phi \, d\phi \, d\theta \tag{2.28}$$

$$S = R^2 \int_0^{2\pi} d\theta \int_0^{\pi} sin\phi \, d\phi = 4\pi R^2 \tag{2.29}$$

If there is a total energy, E_T, spread over the entire surface, S, such that the energy per unit area $\rho = E_T/4\pi R^2$, then there is an an energy, dE, spread over the surface element dS. The energy per unit area, ρ, is then:

$$\rho = \frac{dE}{R^2 sin\phi \, d\phi \, d\theta} \tag{2.30}$$

This is the well known inverse square law for a surface element, $\rho \propto 1/R^2$. Converting to amplitude gives $A \propto 1/R$. Conversion to amplitude is performed because the signal intensity is related to the amplitude by $I \propto A^2$ [75] . The signal intensity is a measure of the time-averaged energy flux, equal to the energy density (energy per unit volume) multiplied by the velocity at which the energy is moving (units $\mathrm{W/m^2}$), and energy must be conserved. The intensity of the acoustic power, I, is related to the angular frequency, ω, acoustic impedance, z, and particle displacement, ξ, by [75]:

$$I = \frac{1}{2} z \omega^2 \xi^2 \tag{2.31}$$

The sound pressure, p, is given by [75]:

$$p = Z\omega\xi \tag{2.32}$$

For a Rayleigh wave generated by a uniform omni-directional emitter, as the waves are confined to the surface, the energy spreads out on the surface of a circle [79]. A line element, dL, on a circle of constant radius, R, is (using polar coordinates, θ is azimuth):

$$dL = R\,d\theta \tag{2.33}$$

$$L = 2\pi R \tag{2.34}$$

As before, if there is a total energy E_T, spread over the entire circumference, L, such that the energy per unit area $\rho = E_T/2\pi R$, then there is an an energy, dE, spread over the surface element dL. The energy per unit area, ρ, is then:

$$\rho = \frac{dE}{R\,d\theta} \tag{2.35}$$

In this case, $\rho \propto 1/R$, and converting to amplitude gives $A \propto 1/\sqrt{R}$. This applies to Rayleigh waves, but for surface skimming compression and shear waves a different relationship has been reported, $\rho \propto 1/R^4$, converting to $A \propto 1/R^2$ [6].

The Rayleigh wave case also applies to bulk waves expanding in a two-dimensional way, such as for a uniform omni-directional infinite line source. In this two dimensional case for bulk waves, the Rayleigh waves would be travelling effectively one-dimensionally. There would be no spreading out of the wave, and hence no reduction in energy due to geometric attenuation.

The shift to the two-dimensional case highlights the importance of the shape, and indeed the directivity, of the source. The directivity gives information on how, if the transducer were a point source, the energy would be distributed with regards to ϕ and θ (or just θ if a line source is being considered). The energy will still spread out as described, but the proportionality relationships with radius would only hold for a given ϕ and θ. The same relationship could be applied at any ϕ and θ, but comparing different ϕ and θ directions at a constant R would not provide a level surface in terms of energy. Unfortunately, for arbitrary shape finite sources, the energy does not spread out as described. Consider the simple example of a finite line source. Far from the ends of the line source, the energy will spread out as $\rho \propto 1/R$, but nearer the ends, this will cease to apply, and the spread will gain some character of a point source. The energy expected at any given point will then not only depend on R as for a omni-directional emitter, nor just R, ϕ and θ as for a directional but point-like source, but on the full position of the point relative to the source.

2.1.6 Intrinsic attenuation

Non-ideal materials cause attenuation through scattering and absorption mechanisms in addition to the reduction in detected amplitude due to geometric factors. A homogeneous material can in some cases be thought of as a large number of small inhomogeneous grains of random orientation, such that when viewed as a whole, the material appears homogeneous. However, the grains still have boundaries with differing acoustic impedance, and this causes scattering, although grain boundaries are very weak scatterers [45].

Intrinsic attenuation takes the form of a standard exponential decay. This exponential decay

groups scattering and absorption into a single attenuation, although for the cases of interest to this project, the dominant attenuation mechanism is scattering. The form of the attenuation equation is derived in the standard way, assuming a constant fractional loss of energy, dE/E, per unit path length (with energy proportional to intensity I), and for length dx [3]:

$$\frac{dE}{E} = -2\alpha\, dx = \frac{dI}{I} \tag{2.36}$$

If A is the amplitude, and d is the distance, then $I = I_0$ and $A = A_0$ if $d = 0$, and α is the attenuation coefficient:

$$I = I_0 e^{-2\alpha d} \tag{2.37}$$

$$A = A_0 e^{-\alpha d} \tag{2.38}$$

$$\alpha = \frac{1}{d} \ln\left(\frac{A_0}{A}\right) \tag{2.39}$$

Attenuation coefficients (combined α), such as those in Kaye and Laby [19] or elsewhere [88], are given as Np/m where Np is dimensionless [75] (and non-SI, but commonly used).

2.1.6.1 Absorption

Absorption of wave energy is simply the standard conversion to heat, and as with scattering, it increases with frequency [75]. It is related to, at room temperature and MHz frequencies, dislocation motion, magnetic structure, thermoelastic effects [89], and friction between grain boundaries [3]. Thermoelastic damping is present in almost all materials, and is due to irreversible heat flow from the lack of thermal equilibrium between contracting and expanding parts of the structure. Shear waves show very little thermoelastic damping as there is no volume change associated with a purely shear deformation [89]. Ferromagnetic materials may partly convert sound energy into magnetic energy, with phase changes during the conversion causing amplitude and frequency dependent hysteresis [3]. Within the frequency range 1MHz to 10MHz, dislocation motion and magnetic structure cause an absorption that is a linear function of frequency.

2.1.6.2 Scattering

For grains on the order of a wavelength, the scatter can be considered as a standard interaction with a boundary; the wave is affected by the anisotropy of each grain it encounters. For smaller grains (more likely to be the case for most of the samples tested), the scatter is still present with a third power dependence on grain size; the wave propagates as though the material has some average of the grain properties (and hence appearing as isotropic and homogeneous for randomly orientated grains), the equivalent of Rayleigh scattering in electromagnetics [3, 75, 90]. Intermediate frequencies therefore experience some material averaging dependent on the grain size [90]. Formulae exist to predict attenuation based on frequency, average grain size, linear elastic hysteresis absorption, and mean value of the fractional variation of the elastic constants of a single crystal [91]. Other work [92] has attempted to quantitatively predict the variation of the attenuation coefficient with frequency and scatterer size by splitting the absorption coefficient into absorption (α_a, with dependence on wavelength, λ) and scattering (α_s, with dependence

on wavelength, λ, and scatterer size, D), and then splitting the scattering into different regimes [88, 92, 93]:

$$\alpha = \alpha_a(\lambda) + \alpha_s(\lambda, D) \tag{2.40}$$

$$\alpha_a = C_a/\lambda \tag{2.41}$$

$$\text{Rayleigh:} \quad \lambda \gg D \quad \alpha_r = C_r D^3/\lambda^4 \tag{2.42}$$

$$\text{Stochastic:} \quad \lambda \approx D \quad \alpha_s = C_s D/\lambda^2 \tag{2.43}$$

$$\text{Diffusive:} \quad \lambda \ll D \quad \alpha_d = C_d/D \tag{2.44}$$

Medium constants are C_a, C_r, C_s, and C_d. Mode conversion occurs at the grains, so scattered energy is in the form of compression and shear waves, and the shear wave attenuation term is dominant for either incident wave mode [88]. It is clear that the frequency dependence of the attenuation characterises which of the loss mechanisms is taking place [6], although as there is a distribution of grain size, the exponent will not usually exactly correspond to one particular scattering mechanism, and exponents in the approximate range of 3 to 3.8 have been observed for steel [88]. There can also be a term $\propto 1/\lambda^2$ to account for dislocation damping, particularly strong during cyclic stress approaching fatigue but potentially also present without such loading [88]. Various formulae exist to predict the attenuation function based on the above relationships [88], by, for example, assuming that scatterer size has a power law distribution [92]. A general increase with temperature is often neglected. Unfortunately, the formulae are not trivial to calculate, and would require more information than would usually be available during an online inspection. Consequently, it is unlikely to be better than an approximate calibration using a reference sample. Other workers assume the scattering is dominated by the Rayleigh regime and hence assume a fourth-power law [89]. Samples with high attenuation will change the pulse shape the further they travel, due to the dependence of attenuation on frequency [1].

Absorption can be dealt with by improving the signal strength, but scattering increases the background noise in step with the signal level, and hence sets a limit on the sensitivity of a transducer [75]. Polycrystalline materials at MHz frequencies usually have negligible absorption relative to the scattering [3], and signal averaging will not reduce grain scatter noise [1]. Due to the considerable grain scattering of iron and steel, ultrasonic testing frequencies are generally restricted to below 5MHz; steel particularly will consist of many different grains of differing size and orientation and potentially different crystal types due to the complicated phase diagram of steel [75, 94].

2.1.6.3 Attenuation coefficients

Attenuation coefficients are believed to vary widely due to differences in manufacturing of steel and other materials [75, 94], including the heat treatment [91]. The microstructure of steel can be changed considerably by heat treatment, and different heat transformation products can have different attenuation coefficients for the same grain size [95]. Reduction of the attenuation usually occurs if a cast structure is worked (cold or hot), such as by rolling or extruding, as the larger

grains are destroyed, reducing scattering [75]. The anisotropy of the specimens that occurs from the different processes is important. Surface roughness also causes attenuation by scattering, with a f^2 dependence, which makes attenuation measurements on samples that have not been prepared very difficult [89].

2.2 Defects in metals

Solidifying metals will, unless very carefully cooled, be polycrystalline. They grow large numbers of grains simultaneously at random orientations, meeting at grain boundaries. Polycrystalline materials are used throughout this work, as single crystal bulk metals are very expensive/difficult to grow, unrealistic of the prototypical application of TOFDI, and highly anisotropic. The material of most interest to this study is steel, an alloy of iron and carbon whose characteristics are altered by quenching and tempering. Low carbon steels contain between 0.1% and 1.7% carbon [94].

The term flaw usually refers to an undesirable condition, whereas defect refers to a flaw that is sufficient to reject the sample [3]. Since the threshold for rejection is not a constant, the terms are used interchangeably within this study.

There are various ways in which defects can enter the steel; during billet production, working into a product, and service of the product. Realistic crack-like defects are often elliptical if they arise from growth during service, but the shape is dictated by processes and geometry if produced during manufacture [2]. Surface breaking cracks can be considered especially important due to stress concentration at sample surfaces, as such defects are more likely to grow [6]. Although no specific defect type was considered during this study, some of the more common defects are discussed here.

2.2.1 Production

When casting or hot forging metal, numerous defects will occur. For example, the metal will of course contract as it cools from a liquid to a solid, and shrinkage flaws are cavities caused if this process is somehow prevented. Such flaws are further classified into macro-shrinkage (piping), centre-line shrinkage (filamentary shrinkage), and microshrinkage. An example of a shrinkage flaw is hot tears/cracking. These result from stress developed near the solidification temperature when the metal is still weak (perhaps if shrinkage is prevented by the mould and there is insufficient feed of metal), and have a ragged form. Internal hot tears are caused by stress from progressive solidification leading to different contraction rates between inner and outer parts of the casting [75, 94].

Once the metal has cooled and is completely solid, remaining internal stress from different cooling rates through the metal can cause cold cracking, a problem more likely to occur on large complicated castings. Blow holes and porosity can also occur during solidification if gas bubbles are released during pouring and cooling or if not all gasses are removed from the mould and cores, as well as inclusions from the metal/mould interaction [75, 94]. An example of a void within the bulk of a cast steel billet is shown in figure 2.4.

Often during the production of a component, it will be necessary to use fusion welding to join two similar metals. This creates a heat-affected zone, and this region can contain defects which include; cracks, cavities, solid inclusions (e.g. slag), lack of (side-wall) fusion and penetration, and imperfect shape. Dividing cracking further, there is; solidification (hot) cracking, hydrogen

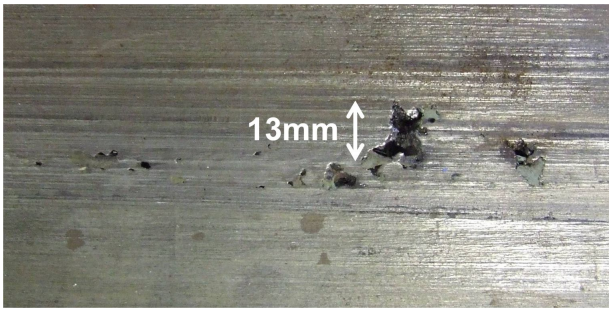


Figure 2.4: An example of a defect within the bulk of a steel billet that has been cut open for inspection. This 13mm void must have been formed during the casting process, as the sample has not undergone any other processes that could have caused such a void to occur.

induced heat affected zone cold cracking, lamellar tearing, and re-heat cracking [1, 6, 94]. Cavities are caused by gas or shrinkage (which may also contain gas as it diffuses into them during or after cavity formation). Gas porosity is spherical pores from a physical or chemical reaction during welding. The main solid inclusions in welds are the non-metallic slag, flux, and oxides, and the metallic tungsten and copper .

2.2.2 Service

During service, cracks can occur for various reasons such as corrosion (known as stress corrosion cracking [37]), but all components under load can suffer from fatigue and creep.

Fatigue refers to the progress of a crack due to a cyclic load (usually below the material yield stress) which causes the component to be in tension for at least part of the stress cycle. The tension causes a crack to extend inwards, often starting at a notch or similar where a stress concentration is present. Crack progression reduces the component cross-section until failure occurs [94].

Creep refers to continuous plastic deformation of a material due to a constant stress under the yield strength but for large time periods. This stress will be within what is usually considered the elastic limit, although not the "true" elastic limit which is much lower, as dislocation movement can occur [94]. Creep is a function of how long the material is stressed for, the material temperature, with the rate of creep being most significant near the melting point (and only important when $T \gtrsim 0.4T_{melting}$), and of course the amount of stress, being most significant near the yield point. Creep is a major issue in industries such as nuclear power and aeronautics [96]. Creep and fatigue acting together is more than a linear combination of the two, and often they are referred to jointly as creep-fatigue [94].

Crack extension requires energy, often provided by the applied external load doing work and from the elastic energy stored in the structure. If all the energy for crack extension is stored elastically in the structure, unstable brittle fracture will result, and no load increase is required for crack propagation.

2.3 Ultrasound interaction with defects

Within wave propagation work, a defect is any discontinuity or boundary of more complex shape than a half-space, including cavities, inclusions, and cracks such as slits and wedges. Scattering occurs due to reflection, diffraction at the edges, and under some conditions, refraction through

the defect [71].

Defect size is assumed to be comparable to the wavelength of the incident ultrasound, as the signal received when the defect size is much smaller than the wavelength will be significantly smaller relative to a signal from a defect of comparable size to the wavelength. These very small defects will then usually be lost in the noise of the B-scan. Some work splits defects into broad classes; weak volumetric scatterers and stress-free boundary defects are often used [37, 38]. For example, casting defects are voluminous and globular [75].

Natural defects differ from artificial reflectors, generally not having regular shapes or smooth surfaces [75]. Defect roughness causes incident ultrasound to have a diffuse, wide-angled component upon interaction [1,97]. A smooth flaw can have a near uniform response when scanned, whereas the combination of specular reflections and scattered/diffracted waves mutually interfering observed with faceted flaws (diffraction from the edges will always be in all directions [1]) can result in small changes in transducer position leading to large changes in the signal returned [30, 75]. This is because signals will only be detected when different parts are sufficiently in phase for constructive interference, requiring the path length to have little variation as the diffraction point varies [1]. Irregular surfaces stop strong specular reflections, but surface roughness only affects reflectivity if it is greater than approximately one tenth of a wavelength [75].

Natural defects can be semi-transparent to ultrasound, but this occurs less in steel than aluminium due to differing oxide properties associated with cracks [75]. Tight fatigue cracks under a compressive force are semi-transparent to ultrasonic examination [13], and these cracks may become more transparent to ultrasound when the load forming the crack is removed [16]. However, crack tips should still provide diffracted signals until the crack transparency is very high [13], although other work reports that crack tips (and grain boundaries) are very weak scatterers [45]. A tightly closed fatigue crack can transmit Rayleigh waves through the crack due to interfacial contact within the crack, with transmission of elastic energy more pronounced near the crack tip, and streams of waves due to scattering by discrete contact surfaces along the crack [98].

Defects can be detected by an extra feature in a B-scan from a scattered wave or from a missing/altered feature compared to a standard scan [51]. For example, if the mechanical properties of a defect vary gradually relative to the wavelength, a signal may not be strongly reflected, and porosity in castings can absorb so strongly that no scattered wave is observable, and hence the defect can only be detected by the shadow on the back-wall [75]. Samples within this work may be thick, such that a strong back-wall reflection cannot necessarily be guaranteed, and variations in sample thickness can make the back-wall reflection difficult to track. Combined with diffraction around small defects, missing features are difficult to observe for all but the largest defects. Consequently, this project focuses on defects which produce a detectable scattered wave.

In this work, representative defects are modelled using finite element method (FEM) computer simulations, sometimes referred to as finite element analysis (FEA). FEM can account for the characteristics of the emitted pulse, an arbitrary scattering defect, and the receiver. FEM is a very broad field, and the exact details of the algorithms used within the FEM software used are not available. PZFlex is a FEM tool used to simulate every important aspect of an ultrasonic testing set-up. The manufacturers claim that the accuracy of the program has been verified by analysing standard benchmark problems for which analytic solutions are known, and that code results have also been compared with those from other computer programs for problems for which analytic solutions are not available. Unfortunately, the makers of PZFlex have not disclosed the source code nor the exact algorithms employed, so it is difficult to say under which conditions the

model will fail. Tests at the University of Warwick have generally shown good agreement, with analytic solutions, other models, and experiment, although there have been exceptions [99]. The conditions PZFlex is being used to model within this work are well understood and have been validated within the research group several times (although a validation has not been published by the group), showing accurate results, consistent with experimental measurements, other models, and analytic solutions. Since the modelling is only required here to provide a test-bed for signal-processing algorithms, and to test concepts before conducting experiments, precise accuracy is less important as the real results will come from the experiment, but the accuracy observed in previous tests mean that results generated are expected to be reliable. PZFlex has also proved useful for qualitatively exploring phenomenon such as wave interactions with defects, mode-conversion, and the appearance of a head wave.

2.4 Electromagnetic acoustic transducers (EMATs)

EMATs can be used to generate and detect ultrasound on metals. EMATs are useful for applications in which the testing must be performed without contact with the sample, or for testing that requires an absence of coupling media and no surface preparations [61–64, 100–105]. Lift-off is typically limited to 1.5mm for single shot pulse-echo measurements [60, 103], and for some configurations, noise has been found to exceed signal at 3mm [106]. However, in work using Rayleigh waves and averaged captures, useful signals have still been observed at 5mm lift-offs [62, 64]. This is possibly due to the lower frequencies used for the Rayleigh wave work, as in addition to being dependent on the EMAT coil design, and the attenuative properties of the sample, the lift-off is also dependent on frequency [107]. In generation, the magnitude of the electromagnetic fields in the material exponentially decrease with lift-off [108, 109]. The lack of contact means that reflected signals are not distorted by additional reflections within the couplant, nor can variations in couplant alter transit times, and energy is not leaked into the transducer; this makes EMATs ideal for transit-time and attenuation measurements [89]. Unfortunately, due to the impedance of the coil, which increases with frequency, they have poor response above a few MHz [6], but this still allows for operation over a fairly broad bandwidth (and hence they can detect temporally sharp pulses). In addition, EMATs can generate waves, such as SH waves, that are difficult to generate with piezoelectric transducers [108]. Although EMATs are non-contact, they are not remote from the surface, and must be cooled [4, 5] or modified [110] for high temperature use.

Typical EMATs consist of a permanent magnet or electromagnet, which provides a static magnetic field (although work has been done exploring pulsed electromagnets [89, 110]) and a coil of wire (figure 2.5). Under some conditions, generation EMATs can operate without the static magnetic field as they generate a dynamic magnetic field [111], but reception EMATs always require a static magnetic field. Generation and reception is via three different mechanisms operating within the skin depth of the sample [74, 89]:

- Lorentz-force from the interaction between eddy currents and the static magnetic flux (there is also an effect due to the interaction with the self-induced dynamic magnetic field, but this is not present in reception);
- Magnetostriction;
- Magnetisation-force between the oscillating magnetic field and the magnetisation of a ferro-

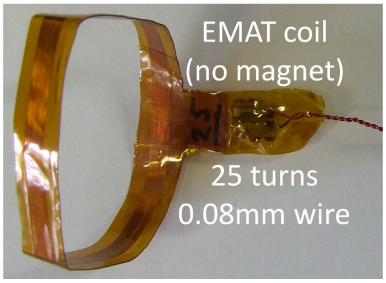


Figure 2.5: This example EMAT coil consists of 25 turns of 0.08mm diameter insulated copper wire between two adhesive layers of Kapton tape. The tape holds and protects the wire. A magnet is placed within the EMAT coil to produce an ultrasound receiver. Thicker wire would be used if the EMAT were for generating ultrasound. The ends of the coil are soldered to a BNC connector (not shown). This example was constructed for basic testing and is not production quality.

magnetic material due to the static magnetic field (does not appear to be present in reception).

EMATs are used as receivers in this project, in preference to a laser interferometer, partly due to the desire for a sparse array of receivers, but also due to the significant problems in using a laser interferometer on the typical rough and dirty (oxide scale) surface of a metal billet in production [6, 7]. Consequently, the following discussion focuses on reception, but aspects of generation are described where relevant. Full descriptions of the mechanisms for generation and reception are available within the literature [74, 89, 112, 113]. The lower conductivity and higher density of steel reduce EMAT efficiency, making ultrasound detection more difficult [102]. EMATs act as velocity sensors [60], and hence will emphasise higher frequencies, since velocity is the differential of displacement, and the differential of an example displacement, $\sin(\omega t)$, is $\omega \cos(\omega t)$, with ω as the angular frequency and t as the time. Note that the descriptions of EMAT operation assume that it can be described sufficiently in classical and macroscopic terms, that the electric, magnetic, and elastic properties of the sample metal are linear, isotropic, and homogeneous, and that the sample can be considered a semi-infinite space [112].

EMATs can detect all wave modes, since the acoustic waves do not travel through a liquid or gas couplant before being detected. However, as described in the section on Lorentz force (section 2.4.2), they only detect motion in a direction perpendicular to the static field direction. Since a given wave will usually arrive at an angle, it will have components of motion in multiple directions. Consequently, EMATs are able to detect all wave modes except at very specific angles. Some applications place the EMAT coil at the edge of a magnet, where the in-plane and out-of-plane components are comparable, and hence both directions of velocity are detected [4]. There is also work describing how Rayleigh waves are generated when the spacing of a meander coil is equal to an integer number of wavelengths of Rayleigh waves ($n\lambda = 2L$ with wavelength λ and spacing L), and bulk waves are generated at specific angles and wavelengths ($n\lambda = 2L \sin\theta$ with angle from the surface normal θ), again dependent on the meander line spacing, meaning that the bulk waves have a frequency which is angularly dependent [114].

EMAT receivers generally produce quite small peak-to-peak voltages (typically in the range 500µV to 5µV, although sometimes as high as 2mV), and hence can require high gain high impedance amplifiers with low noise just to bring the signals up to a level a typical analog-to-digital converter (ADC) can detect [109].

2.4.1 Skin depth

Consider an alternating voltage applied to a large conducting wire, producing an alternating current (AC) in the conductor. The AC density, i, in the surface of a conductor decreases with depth, z (m), and is also a function of frequency, f (Hz), magnetic permeability, $\mu = \mu_0\mu_r$ (H/m) with magnetic permeability of free space $\mu_0 = 4\pi \times 10^{-7}$ H/m and relative permeability μ_r (dimensionless), and electrical conductivity, $\sigma = 1/\rho$ (S/m), with resistivity ρ (Ωm) [115]:

$$i = i_0 e^{-z\sqrt{2j\pi f\mu\sigma}} \tag{2.45}$$

The current at the very surface is i_0. Of course, from evaluation of $j = (a+bj)^2$, with $j = \sqrt{-1}$, and a and b as real numbers:

$$a = b = 1/\sqrt{2} \tag{2.46}$$

$$\sqrt{j} = 1/\sqrt{2} + j/\sqrt{2} \tag{2.47}$$

$$i = i_0 \left(e^{-z\sqrt{\pi f\mu\sigma}} e^{-jz\sqrt{\pi f\mu\sigma}}\right) \tag{2.48}$$

The skin effect is the phenomena in which high-frequency AC and magnetic flux only penetrate to a limited depth of penetration known as the skin depth, δ, at which the field is $1/e$ of its original value [115, 116]. If the thickness of the conductor is much greater than this depth, high frequency alternating currents are considered to be confined to the surface rather than the volume of the conductor [115]. Defining the penetration depth, δ, as [74, 115–117]:

$$\delta = 1/\sqrt{\pi f\mu\sigma} \tag{2.49}$$

$$i = i_0 \left(e^{-z/\delta} e^{-jz/\delta}\right) \tag{2.50}$$

The current density has a real term, which is the commonly used term for how the current density decreases with depth (figure 2.6), and a phase term. There is a phase lag equal to z/δ at a depth z [116], relative to the phase of the AC on the very surface of the conductor. Assume now that the conductor can be considered flat, but without edge effects, and has length (a dimension parallel to the direction the AC is passed and in the surface plane), width (a dimension perpendicular to the direction the AC is passed and in the surface plane), and depth (a dimension perpendicular to the direction the AC is passed and perpendicular to the surface plane). Integrating for infinite depth over a section of the conductor of width w, gives the total current, I, passing through the conductor [115]:

$$I = w\int_0^\infty i\,dz = i_0 w\delta/(1+j) = i_0 w\delta e^{-j\pi/4}/\sqrt{2} \tag{2.51}$$

The total current magnitude is $i_0 w\delta/\sqrt{2}$ with a phase shift of $-\pi/4$ relative to the phase of the AC on the very surface of the conductor. This assumes that the conductor is much deeper than the penetration depth. For an aluminium sample with resistivity of 24.2nΩ m (at a temperature of 273.2K) [118], $\mu = \mu_0$, and a frequency of 100kHz (the lowest frequency that is likely to be generated in this work), the skin depth is 0.25mm. At this depth, the AC would still have 37% of

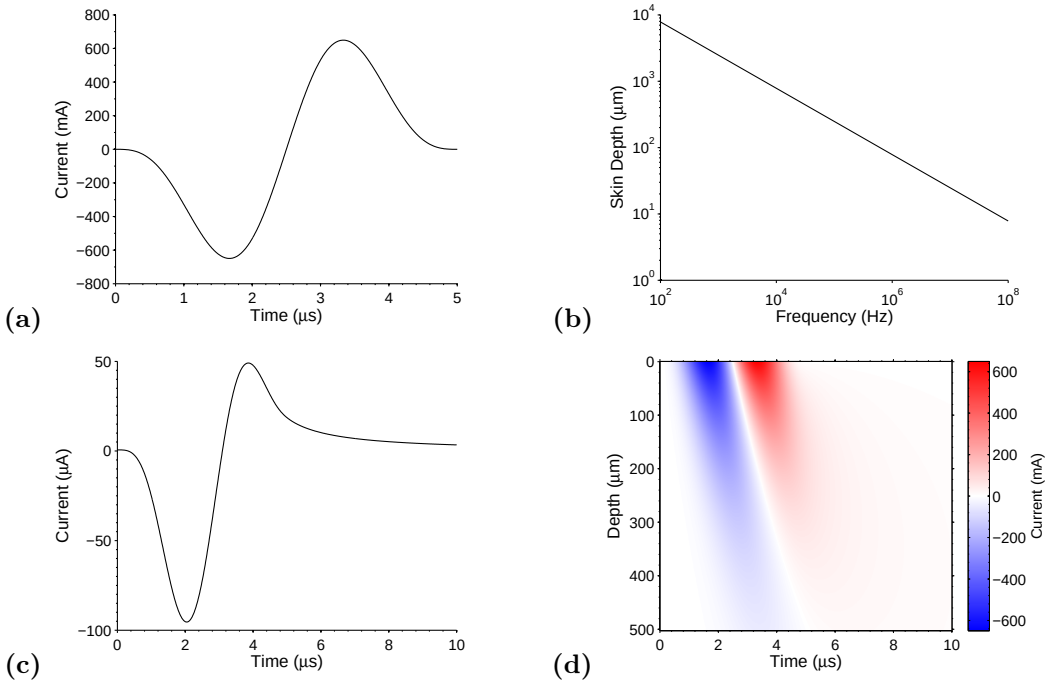


Figure 2.6: Consider a current pulse (a) consisting of a 200kHz sine wave that has been Hann windowed, applied to a large aluminium conductor. The first minima either side of the centre frequency are at 0kHz and 600kHz. In an infinite depth aluminium sample of unity width, the total current is attenuated in the frequency domain as defined by the skin depth (b) with a factor of $1/\sqrt{2}$, and phase shifted by -45° relative to the AC on the very surface of the conductor (c). The lower frequency parts of the signal penetrate deeper (d), which subtly changes the shape of the signal in the time domain from that seen in (a) at the surface, only forming (c) when integrating over infinite depth.

its energy. After $5\delta = 1.25$mm the AC only has 0.7% of its energy, and hence it could only really be considered as confined to the surface if the depth were at least 1.25mm (i.e. this is an effective infinity).

The voltage, V, on the surface along the length, l, of the conductor is obtained from (the length is perpendicular to the width, w) [115]:

$$V = i_0 l \rho \tag{2.52}$$

The current density is related to the electric field in the medium, E, by [115, 117]:

$$i = \sigma E \tag{2.53}$$

There is a more full equation for the skin depth, which includes conductors which cannot be considered good, and includes terms for electric permittivity ε (F/m), and angular frequency $\omega = 2\pi f$ (radians/s) [116, 117]:

$$\delta = \frac{1}{\omega\sqrt{\varepsilon\mu_0\mu_r\left(\sqrt{1+(1/\rho\omega\varepsilon)^2}-1\right)/2}} \tag{2.54}$$

Considering the good conductor that is mild steel, the values can take on a significant range depending on the exact composition, the temperature, field strength, and frequency. However,

taking some example values, the resistivity of mild steel is 170.2nΩ m (at a temperature of 273.2K) [118], and the relative permeability can be estimated to be between 1 (more likely for higher temperatures, fields, and frequencies) and 1000. Using $\mu_r = 1$ (to produce a larger skin depth), it can still be seen that any effects can be considered limited to the surface, as for frequencies as low as 100kHz the skin depth is approximately 650μm and therefore $5\delta \sim 3.3$mm. For $\mu_r = 1000$ the skin depth would be approximately 32 ($\sqrt{1000}$) times smaller again (approximately 21μm) but would otherwise be identical. The results are applicable to plane boundaries, and are applicable to curved boundaries when the radius of curvature is at least five times larger than the skin depth [116]. Frequencies of interest to this work are typically in the low MHz range.

It is the skin depth that allows the current in the conductor to be treated as an image of the current through the wires of the EMAT coil. The current passing through the EMAT coil creates a magnetic field around the conducting wires of the coil. The current is AC (as it is a pulse), producing a varying magnetic field which is incident upon the conducting sample. Eddy currents are created when a conductor experiences a change in the magnetic field (in this case because the magnetic field is changing whilst the conductor is stationary), as defined by Faraday's law of induction, which states that the electromotive force (EMF) generated is proportional to the rate of change of the magnetic flux. However, the incident field from the EMAT coil will only penetrate to the skin depth, and therefore the eddy currents will resemble an image of the wires if considered over the depth of penetration (as opposed to if the eddy currents were still found deep within the conductor), with the appropriate change in phase and amplitude. In work that used EMATs for ultrasound generation of Lamb waves for the purposes of texture measurement, the drive current pulses were measured, and they appear similar to half a period of a sine wave, with the duration being 5-6μs (making an entire period 10-12μs) [102]. If over the region necessary for the sample to be considered infinite, there was a static magnetic field, then the force generated could be calculated using the Lorentz force equation.

2.4.2 Lorentz force

The Lorentz force mechanism is due to the interaction between eddy currents and the static magnetic flux density [75, 89]. The Lorentz force, $\mathbf{F}$, is given by the standard equation including the electric field, $\mathbf{E}$, the magnetic flux density, $\mathbf{B}$, and the velocity of the particle, $\mathbf{v}$, with charge q:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.55}$$

When an ultrasound generation EMAT is placed close to a conducting material, and an RF current is passed through the coils, an eddy current is induced in the surface region of the sample which interacts with the static magnetic field from the magnet [107], with the eddy current penetration depth discussed in the section on skin depth (section 2.4.1). The resultant Lorentz force acts upon the electrons carrying the eddy current [107], as they can be considered "free" relative to the fixed positions of the ions in the lattice. The cross term means that the Lorentz stress generated is perpendicular to the current ($q\mathbf{v}$, from the charge on the current carriers, q, and the mean velocity of the charge carriers, $\mathbf{v}$) and the static biasing magnetic flux density, $\mathbf{B}$. Writing the Lorentz force in terms of the eddy current, $\mathbf{J}$ [112, 119]:

$$\mathbf{F} = \mathbf{J} \times \mathbf{B} \tag{2.56}$$

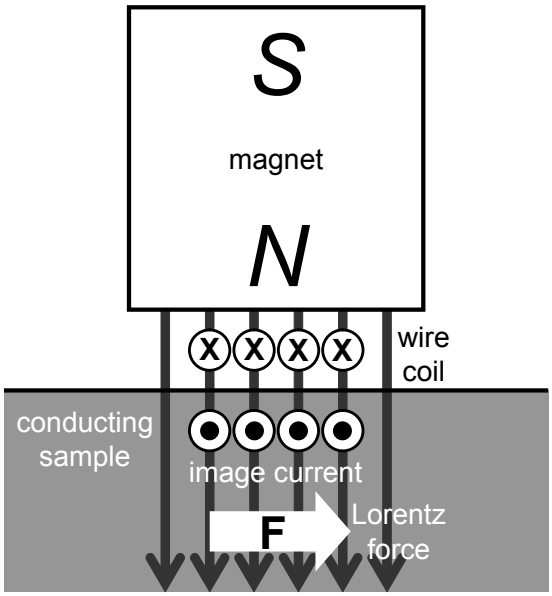


Figure 2.7: A simplified diagram of an EMAT. An RF current is passed through the EMAT wire coil, inducing an eddy current in the surface region of the sample. This interacts with the static magnetic field from the EMAT magnet, to produce a Lorentz force, and consequently an ultrasonic vibration is generated.

By colliding with the ions in the conductor (and other transformation processes), the Lorentz force generates an ultrasonic vibration [107,120]. The coupled Maxwell and elastic wave equations are available in the literature [119, 121]. The form of the force generated is described in the literature [119], but is not required here due to the focus on detection. Figure 2.7 is a simplified diagram showing the direction of the force generated.

The Lorentz force technique is reversible and hence an EMAT can generate and receive ultrasound of the same mode [74, 75, 121, 122]. Due to an ultrasonic wave propagating through a material, the atoms (consisting of a positive ion and electrons) of the solid will move in the same direction as the oscillation and in phase with it. When the ions and associated electrons move in the presence of a static biasing magnetic field, the electrons are affected by the Lorentz force to a much greater extent due to the difference in mass between the two particle types ($\mathbf{F} = m\mathbf{a}$ with mass m and acceleration $\mathbf{a}$), allowing the heavy ions to be neglected when considering reception. The dynamic electric field, $\mathbf{E_d}$, induced by the deformation (instantaneous ion velocity $\mathbf{v}$) and the biasing magnetic flux density, $\mathbf{B}$, in a conducting material is (from the Maxwell equations) [89]:

$$\mathbf{E_d} = \mathbf{v} \times \mathbf{B} \tag{2.57}$$

If the ultrasonic wave were uniform within the region of the sample being considered, instantaneous ion velocity $\mathbf{v}$ will be the same throughout this region. If the biasing magnetic flux density $\mathbf{B}$ was also constant over this region, then the induced dynamic electric field $\mathbf{E_d}$ would also be constant. Ohm's law (in one possible form) gives the current density, $\mathbf{J}$, for an applied electric field, $\mathbf{E}$, and conductivity, σ [89], at the surface [60]:

$$\mathbf{J} = \sigma\mathbf{E} \tag{2.58}$$

This is only valid for a static conducting material taken as the reference frame. Including the motion of the conductor relative to a static magnetic field [89]:

$$\mathbf{J} = \sigma \left(\mathbf{E} + \mathbf{v} \times \mathbf{B} \right) \tag{2.59}$$

$$\mathbf{J} = \sigma \left(\mathbf{E} + \mathbf{E_d} \right) \tag{2.60}$$

The coil in the EMAT detects this dynamic electromagnetic field [107]. It has already been stated that an EMAT is a velocity sensor [60], and this can be seen in the above equations. Stating the above in terms of the acoustic displacement field, **S**, involves taking the differential, which for a sinusoidal excitation effectively just produces a $\partial/\partial t = j\omega$ term [121] (noting that dimensionally, **S** will now have units of m/s rather than m as it had prior to differentiation):

$$\mathbf{J} = \sigma \left(\mathbf{E} + j\omega \mathbf{S} \times \mathbf{B} \right) \tag{2.61}$$

Consequently, a current is produced in the EMAT coil due to the motion of the conducting sample in the presence of the static magnetic field. Reception requires a static magnetic field, whereas generation can also use the dynamic magnetic field of the pulsed wire coil [74, 89]. Of course the electromagnetic field can only penetrate a short distance, as discussed with regards to skin depth, and so only ultrasonic waves very close to the surface (within a few skin depths, with only approximately 37% of the electromagnetic wave energy remaining after one skin depth) can be detected. When it was assumed that the ultrasonic wave and magnetic field were approximately uniform within the region of the sample being considered, this region is only the same region very close to the surface; if they are not uniform, they will still be detected in some form, this assumption was just made to ease the discussion of EMAT reception of ultrasonic waves.

It is often desirable to optimise for one application. Coil inductance, both self and mutual, affects the signal generated or received. Mutual inductance is dependent on frequency, the electromagnetic characteristics of the sample, coil configuration including shape, size and number of turns, and the lift-off. Coil configuration also determines the self-inductance and resistance (the latter of which is also dependent to some extent on wire thickness).

2.4.3 Magnetostriction

Magnetostriction is mechanical deformation of a material placed in a magnetic field and can be observed in most ferromagnetic metals [3, 74, 75, 123, 124]. For non-ferromagnetic metals, there is a linear relationship between the applied static magnetic field and the amplitude of the generated wave, but this is not necessarily observed in ferromagnetic metals due to magnetostrictive effects [125]. It is often claimed that magnetostriction will dominate over the Lorentz force in ferromagnetic metals, regardless of the bias field direction [89,106,119,122]. However, other research, using theory and experiment, suggests that when generating low frequency Lamb waves, in iron and some steels, with the bias field direction parallel to the surface plane (the same direction as the propagation), that the Lorentz force will dominate for strong magnetic fields, and that the magnetostriction will dominate for weak magnetic fields [113,125,126]. The Lorentz force vanishes if the magnetic bias is parallel to the coil elements, but significant magnetostrictive generation can occur [113]. When the bias is normal to the surface, the magnetostriction is quite small; the Lorentz force mechanism dominates the sensitivity in the near normal direction, but the magnet-

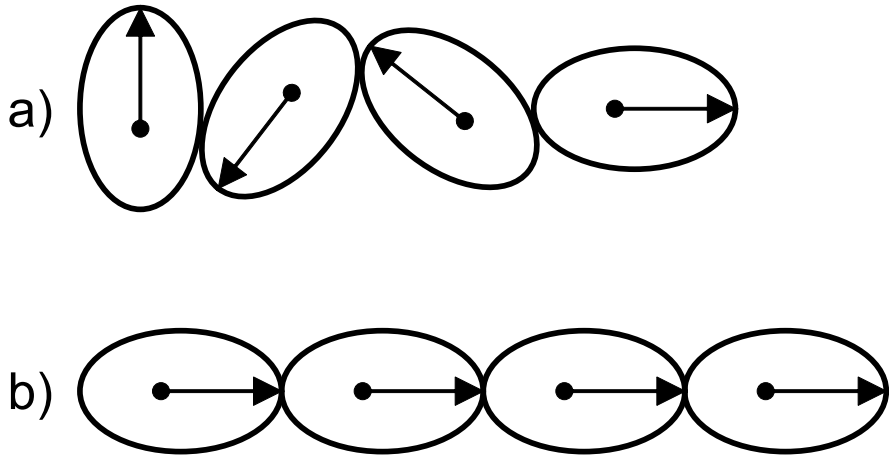


Figure 2.8: The black dots represent atomic nuclei, the arrows are the net spin moment of each atom, and the ellipses are the electron cloud. The cloud is assumed to be elongated from spherical along the direction of the spin moment. The upper row (a) represents the disordered paramagnetic regime. Below the Curie temperature, the spins become ordered in the ferromagnetic regime, causing spontaneous magnetisation, as shown in the lower row (b). The turning of the spins drags the electron clouds with them. In reality, the spin-orbit coupling is weak, and hence this diagram is hugely exaggerated [123].

isation mechanism contributes significantly to the sensitivity at higher angles, and this has been confirmed theoretically and experimentally [113]. Recent research, using theory and experiment, demonstrates that for a bias field normal to the surface plane, the Lorentz force is the dominant transduction mechanism on steel materials [127]. It is EMATs with a bias field normal to the surface plane, operating on steel, which are of interest to this project, and hence it can be considered that the Lorentz force is the dominant transduction mechanism. However, a brief discussion on magnetostriction is still provided in this section.

Spontaneous magnetostriction of ferromagnets occurs due to the ordering of magnetic moments into small regions, called domains, at the Curie temperature, each domain being spontaneously magnetised, or self-saturated, with its spins almost completely aligned, depending on the temperature; the directions of magnetisation of different domains need not be parallel [123, 128]. The increase in the gross magnetic moment of a ferromagnetic specimen in an applied magnetic field takes place by two independent processes. In weak applied fields, the volume of domains favourably orientated with respect to the field increases at the expense of unfavourably orientated domains. In strong applied fields, the domain magnetisation rotates towards the direction of the field [128]. This results in the length, l, of a ferromagnet, changing (δl) with applied magnetic field, in a process known as field-induced magnetostriction. However, the changes in length are very structure sensitive, so no general formula relating to the two is possible [123, 124]; for example, the bulk magnetostriction of iron changes sign as the magnetic field is increased [123, 124]. At first, iron (and some steels) lengthen with applied field, and then from that point shorten again, even becoming shorter than their initial length at some field strengths, but with little volume change at any point [126]. Magnetostriction curves for various polycrystalline materials, including iron, are available in the literature, with the magnetostriction for iron and (some) steels above saturation being small [126]. Ultimately, these deformations are due to spin-orbit coupling (figure 2.8), leading to a change in lattice constants as the atomic moments are reorientated [113, 123, 128].

For an ideal case of a magnetic field applied perpendicular to the easy axis (the energetically

favourable direction of the spontaneous magnetisation in a ferromagnetic material) in a single crystal with uniaxial anisotropy, there is a straightforward equation [124]:

$$\lambda = \frac{3}{2}\left(\frac{|\mathbf{M}|}{|\mathbf{M_s}|}\right)^2 \lambda_s \tag{2.62}$$

In this case the magnetostriction ($\lambda = \delta l/l$) is a function of the magnetisation, $\mathbf{M}$, the saturation magnetisation, $\mathbf{M_s}$, and the saturation magnetostriction, λ_s. This is known as Joule magnetostriction, the direct magnetostrictive effect [122,126]. Due to the square term, and unlike the Lorentz force, the magnetostriction is not dependent on the sign of the magnetisation, only its magnitude [3, 75]. Consequently, if the magnetic field changes sign, there is a discontinuity in the time differential of the magnetostrictive force, as it is the absolute value of the magnetic field that determines the magnetostriction, not the direction. A biasing magnetic field can be applied to ensure the varying magnetic field does not change sign, or if available, a strong enough biasing field can be applied to saturate the surface, reducing the possibility of a time varying field changing the overall magnetic field sufficiently to cause any change in the magnetostrictive force [3,74]. This would result in no magnetostriction being observed, but the standard case is simply a biasing but not saturating magnetic field. However, as has already been mentioned, the sign of the magnetostriction can change due to the magnitude of the applied field in more realistic cases, but again, not due to the direction of the applied field. This is also in contrast with the piezomagnetic effect which can change sign due to the magnetic field direction, although some work [89, 108, 122] has erroneously linked the two, probably as they appear similar with a sufficiently strong biasing field (such work often states that it only takes a form similar to that of a piezomagnetic material [108], or that it is just a convenient analogy [122]). In fact, there are no known ferromagnetic or ferrimagnetic materials that are piezomagnetic [123]. Although an already partially magnetised sample can have its magnetisation changed by stress alone (an effect called the inverse magnetostrictive effect), this does not mean that if initially demagnetised the material can become magnetised by stress alone in the absence of a magnetic field [123]. Since the dynamic change of the magnetic field is usually small compared to the large static magnetic field, the analogy can usually be considered valid with an appropriate biasing field, and in addition, the magnetostriction curve can (under some conditions) be considered linear over such a small range, even though in general it is highly nonlinear [122].

There is a transverse magnetostriction accompanying the usual longitudinal magnetostriction [124]:

$$\lambda_t = -\lambda/2 \tag{2.63}$$

Deformation at constant volume is called linear magnetostriction (occurring only below the Curie point), volumetric magnetostriction involves volume change (occurring both below and above the Curie point) [75]. Most materials have a very small volume magnetostriction (or are isovolume such as for a nontextured polycrystalline ferromagnetic material due to randomly oriented easy axes averaging out the anisotropic magnetostriction of individual magnetic domains [89, 119]), and hence magnetostrictive effects are not normally observed above the Curie temperature of the sample [54]. There is no linear magnetostriction in the demagnetised state, and for most materials, magnetostriction is essentially equi-volumetric with little dimensional change occurring after saturation [126]. Hysteresis is usually neglected [122].

The force, **F**, due to the magnetostriction force, is a function of the magnetising field, **H**, and the tensor of the magnetostriction constants, w, which contains material constants that are dependent on the magnetic state of the material [113, 119] (in this case, as w is a matrix, $w \cdot \mathbf{H}$ is equivalent to matrix multiplication):

$$\mathbf{F} = \partial(w \cdot \mathbf{H})/\partial x_i \tag{2.64}$$

Due to the skin depth of the time varying magnetic field, magnetostriction is a surface phenomenon in metals, which means that as an ultrasound source it has a similar shape to the Lorentz force source, but compressional waves are more likely to be generated than the shear waves from the Lorentz force [74]. The form of the the force generated is described in the literature [119], but is not required here due to the focus on detection. Whereas the Lorentz force is directly proportional to the magnetic field, and will continue to increase after the material is magnetically saturated, magnetostrictive generation can be expected to be proportional to the changes in magnetostrictive stress produced by the alternation of the dynamic field about the static bias (if linearity of the magnetostriction curve is assumed, although this is not generally the case) [126].

Magnetostriction is not directly reversible (although apparently reciprocity holds for linearised magnetostriction [122]), but elastic deformation from ultrasonic waves changes the magnetisation state (as it alters the magnetic-domain volume or electron's sub-shell orbit [89], known as inverse magnetostriction [119] or the Villari effect [3, 122]) which will change the flux density of an applied magnetic field, and then a voltage will be induced in the receiving coil. As with the Lorentz force, this means that a static magnetic field is required for detection, whilst it is only optional for generation [74, 75, 89]. In the presence of a polarising magnetic field, the flux density varies linearly with the applied stress [3]. Relating the biasing magnetic field, H_j, and strain, S_J, to the magnetic flux density, B_i, via the normalised permeability tensor, μ^S_{ij}, and coefficients relating the induced magnetic flux density to the applied strain field, $\widetilde{e}^{(MS)}_{iJ}$ [89]:

$$B_i = \mu_0 \mu^S_{ij} H_j + \widetilde{e}^{(MS)}_{iJ} S_J \tag{2.65}$$

Summation is over the axes (matrix columns, $i, j = x, y, z$) and matrix rows ($J = 1, 2, \ldots, 6$). This quantity will vary with angle of incidence of an ultrasonic wave.

2.4.4 Other generation effects

Other generation effects include the magnetisation force (not the same as the magnetostriction force), which is a function of the magnetising field, **H** (with each vector direction differentiated spatially in that direction within the equation), and the bias magnetisation, $\mathbf{M_0}$. Integrated over the volume of the sample being acted on, and including the magnetic permeability of free space $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H/m}$ [89, 119, 125]:

$$\mathbf{F} = \mu_0 \int (\mathbf{M_0} \cdot \frac{\partial \mathbf{H}}{\partial x_i})\, dV \tag{2.66}$$

The magnetisation force is caused by the reorientation of magnetic dipoles due to the interaction between the dynamic magnetic field and the bias magnetisation [129]. For the sort of materials of interest to this work, this can be neglected for a biasing field normal to the sample surface, as magnetostrictive forces are significantly larger than magnetic forces when Lorentz forces are small, and in the high-field limit, the Lorentz forces are dominant [125].

There is also a traction force normal to the surface per unit area, T_{33}, due to the bias magnetisation (M_n is the normal component at the surface of the magnetisation of the metal) having a steep change near the surface of a magnetised metal (this force disappears inside the material) [89, 119]:

$$T_{33} = M_{03} M_n / \mu_0 \tag{2.67}$$

2.4.5 Lift-off

Lift-off has a strong effect on the eddy current, which in turn has an effect on the Lorentz force [130]. The effective width of the excitation force will be related to the width and lift-off of the coil, meaning that lift-off causes both a drop in efficiency and a change in the directivity of the source [74]. If the lift-off is very small, small variations in lift-off result in a significant change in efficiency and directivity [106]. If the EMAT is to be scanned above a surface, and the lift-off is expected to vary, a lift-off should be chosen that balances this variation with the SNR [106]. Lift-off has also been observed to change the frequency content of signals emitted or received by EMAT [63, 64]. Large EMATs (with large magnets) have been found to be less sensitive to lift-off in FEM models [131], but in general, this is highly dependent on EMAT coil design.

2.4.6 Magnets

Many EMATs use neodymium-iron-boron ($Nd_2Fe_{14}B$) sintered magnets, a type of rare-earth magnet made from an alloy of rare earth elements; in this case neodymium is a rare earth element, also referred to as a lanthanide element. These magnets have the highest remanent flux density, B_r, of the available permanent magnets [89]. The crystalline structures have high magnetic anisotropy, making such a structure easy to magnetise in one direction, but difficult to magnetise in the other.

The parameters of interest are:

- Remanence B_r (T), the magnetic flux density after the external magnetising field has been removed;
- Coercivity H_c (kA/m), the intensity of the external magnetic field required to reduce the magnet magnetisation to zero after it has been saturated (note that there is a difference between the field required to make the net magnetic flux density zero in the presence of the external field and the field required to make the magnet lose its magnetic polarisation);
- Energy product BH_{max} (kJ/m^3), the density of magnetic energy;
- Curie temperature T_c (K), the temperature at which the magnet becomes (reversibly) paramagnetic.

For example, a N48 graded magnet has $B_r = 1.4$T, $H_c = 900$kA/m, and $BH_{max} = 370\text{kJ}/m^3$ (typical values). The magnets have a low Curie point of 610-630K [124], making the magnetic properties very sensitive to temperature. Consequently, if they are to be used in high temperature applications, they must be cooled.

Unfortunately, the magnets are electrically conductive, and in generation, ultrasound fields are produced both in the sample and the magnet unless precautions are taken. Roughening the surface causes randomised reflections within the magnet, minimising coherent interference with the actual signal in the sample [89]. It is more usual though, to electrically screen the magnet, using a copper shim for example.

2.5 Laser generation

Unlike with piezoelectric transducers or EMATs, laser beams can either generate or receive (via interferometry) ultrasound [74], but not both with the same laser beam. This work is only concerned with using laser beams to generate ultrasound. The advantages of the laser within NDT include being able to have a relatively large distance between the laser and the sample, and temporally sharp pulses for a pulsed laser system (where temporally sharp laser pulses for this application area can generally be defined as having a duration less than 100ns), that can be formed into a spatially sharp line using a cylindrical lens [50, 59, 73]. Pulsed lasers are suitable for ultrasound generation due to the high radiance they offer; a high power level, emitted over a small solid angle, leads to a high irradiance (power per unit area) [6, 132]. Lasers can generate bulk waves and SAWs, with Rayleigh waves generally having the largest amplitude of the SAWs for the thick samples used within this study [6].

Lasers can generate short duration (high bandwidth) ultrasound thermoelastically (for lower power densities) or ablatively (for higher power densities), but there is actually a continuous spectrum of intermediate regimes with a combination of thermoelastic and ablative properties [6]. This work is more concerned with ablative generation as it generally provides much higher amplitude ultrasound [74], but a description of thermoelastic generation is given for contrast. The energy delivered is easily controllable with optical filters (such as neutral density filters). There are issues to consider regarding the effect of a high sample temperature when using a laser to generate, such as a lower threshold for ablation [6], but since no tests were actually conducted using raised temperatures, it is not covered here.

Not covered in this section is the air-breakdown source, which occurs when a laser pulse is focused to a point just above the specimen, causing the air to ionise and form a plasma. The resultant detonation wave and expanding plasma impacts the sample and causes a normal reactive force which creates ultrasound, resembling the ultrasound generated by an ablation source, but without the surface damage [6, 133].

2.5.1 Pulsed operation

The laser used for this study operates in a Q-switched mode. Q characterises the resonant cavity's ability to store radiant energy, large values indicating good storage, low values indicating the energy will emerge rapidly. For a laser cavity containing light at angular frequency, ω (moving at speed c), with a perfect reflecting mirror at a distance, D, from an output mirror with reflectivity, R [132]:

$$Q = \frac{2D\omega}{(1-R)c} \tag{2.68}$$

In Q-switching, the laser medium is pumped to a highly excited state, and then switches from a high to a low Q, releasing a large amount of energy in a small amount of time (1-100ns) [6, 132]. The energy is released in a pulse of higher peak power and shorter duration than would be achieved without Q-switching. The laser used for this study employs the common Pockels cell method, in which a KH_2PO_4 crystal rotates the plane of the polarisation of the light when an electric field is applied, forming a rapid optical switch in conjunction with a polariser [6].

Cavity dumping, in which the laser is excited continuously with a resonant cavity of high Q and pulses are periodically switched out of the cavity [132], can also be used in pulsed operation, but is not employed within this work as the high pulse repetition rates it can achieve are not required.

In contrast, normal pulsed mode is little more than rapidly switching the power on and off [6,132], and cannot produce the desired pulse duration (more on the order of 100μs) or power. The laser medium is brought to its excited state by, for example, the flash lamp, and once it has reached that state, energy is released in a pulse. The pulse ends as the medium has released sufficient energy that it is no longer in the excited state required for lasing. However, it is likely the flash lamp is still providing energy, and hence the process may repeat. Consequently, a large number of low energy pulses may be observed, each typically longer in duration but lower in power than the equivalent Q-switched pulse. The laser available can be operated in this mode (by switching off the Pockels cell), but the power has been observed to be not even sufficient for thermoelastic generation.

2.5.2 Skin depth and heat diffusion

If a laser beam is incident on a metal surface, the EM fields induce currents in the conduction band. Most of the energy is reflected while the rest is absorbed resistively (e.g. electron-phonon scattering processes) and converted to heat. This process is confined to a penetration depth, δ, by the screening effect of the conduction electrons [3,6,74], as described in the EMAT section (section 2.4) by equation 2.49. However, this formula is invalid at wavelengths much below the 1064nm of a Nd:YAG due to quantum mechanical effects [6, 74]. Skin depths for steel and aluminium are typically a few nanometres; for example, using a wavelength of 1064nm and the properties of aluminium ($\mu_r = 1$, $\sigma = 4 \times 10^7 \Omega^{-1} m^{-1}$ [118]), the penetration depth is 5nm.

Heat flow depends on the thermal conductivity, K. This is defined as the quantity of heat transmitted, Q, for a unit temperature (T) gradient in unit time under steady conditions in a direction normal to a surface of unit area (heat transfer is then only dependent on the temperature gradient) [134]:

$$K = {}^{-Q}/_{\frac{\partial T}{\partial n}} \tag{2.69}$$

Aluminium has $K = 236\text{W m}^{-1}\text{K}^{-1}$ at 273.2K (thermal conductivities vary over the temperature range) in a pure polycrystalline form (thermal conductivities are lower for less pure samples) [134]. Steel has a far lower thermal conductivity, depending on its composition; for example, a carbon steel will typically have $K \sim 50\text{W m}^{-1}\text{K}^{-1}$ at 273K [134]. The "thermal diffusivity" $k = K/\rho c_P$ is often used for heat flow, including specific heat (c_P) and density (ρ), and it determines the speed at which a material will accept and conduct thermal energy [132]. At 273K, the specific heat of aluminium and low carbon steel are $c_P = 880\text{J kg}^{-1}\text{K}^{-1}$ and $c_P = (440-450)\text{J kg}^{-1}\text{K}^{-1}$ respectively [135], and they have densities of $\rho = 2698\text{kg/m}^3$ [136] and $\rho = 7900\text{kg/m}^3$ [137] respectively. The corresponding diffusivities are $99\mu\text{m}^2/\text{s}$ for pure aluminium and $14\mu\text{m}^2/\text{s}$ for mild steel, although other values place commercially pure aluminium at $85.0\mu\text{m}^2/\text{s}$ and carbon steel (1.22 C, 0.35 Mn) at $11.9\mu\text{m}^2/\text{s}$ [132]. In either case, heat travels approximately seven times faster in aluminium than carbon steel.

The heat energy deposited by a pulsed laser starts to diffuse immediately by thermal conduction, and thermal conduction will occur even during the short time during which a pulsed laser beam is impinging on a surface. This means that the energy required to ablate the surface is much greater than might be predicted based on the skin depth alone [6, 74]. The depth of penetration of heat (D) in pulse duration time t is approximately [132]:

$$D = 2\sqrt{kt} \tag{2.70}$$

The laser used within this study has a stated pulse duration of 8ns, leading to a depth of penetration (during the blast period) of 1.6μm for aluminium and 0.6μm for carbon steel. The depth in which the temperature rise occurs is still much smaller than the ultrasound wavelength and hence it can still be considered a surface phenomenon; the rise in temperature is still proportional to the energy deposited by the laser [74]. However, the higher output of a Q-switched laser may mean that due to the high irradiance, the surface is vapourised before conduction can occur, so the quantity vapourised is primarily dependent on the latent heat of vapourisation, L; this occurs above a certain power density, I_c ($I_c \sim 10^{11}$W m^{-2}) [6]:

$$I_c \geq 2L\rho\sqrt{k/t} \tag{2.71}$$

For the laser used, considering a duration of 8ns and a energy delivered of 160mJ (measured), focused to a line (20mm long, 2mm wide, for an area of 40μm^2), the irradiance is $I \sim 5\times10^{11}$W m^{-2} during the pulse (assuming a constant pulse output, although peak output is likely to be higher with lower power in-between), hence thermal conduction can largely be ignored. If focused to a point (rather than a line) only 2mm in diameter, then $I \sim 6 \times 10^{12}$W m^{-2}.

The exact shape of the surface and any contaminants present will affect the acoustic energy generated, although there is not thought to be any adverse effects due to variations in surface roughness [74]; a rough or dirty surface can actually enhance the source, whereas a clean and polished surface will absorb the minimum energy [7]. Radiation pressure contributes to the signal generated, but the effect of this is usually several orders of magnitude less than thermoelastic and ablative effects discussed in the following sections [6].

2.5.3 Thermoelastic generation

Thermoelastic laser generation of ultrasound does not damage the material [73]. The laser energy deposited into a restricted volume, V, at the metal surface, causes a rapid increase in temperature, ΔT, and a corresponding (small) thermal expansion to $V + \Delta V$; the volume is free to expand normal to and away from the surface but is constrained from expanding in the plane of the surface by the surrounding material. This in turn causes an elastic strain, ε, and if situated in the bulk, the effect is equivalent to the production of extra volume and would create a spherical compressional wave source. However, the source is actually at the surface, as the energy is being deposited by the laser (and the laser spot size is much larger than the depth to which the heat penetrates), such that little compression wave energy is observed normal to the surface, and in addition, shear waves are also seen due to mode conversion [6, 49, 74, 75]. With u as the displacement along an axis and x as the direction of the axes:

$$\varepsilon_{ii} = \frac{1}{3}\frac{\Delta V}{V} = \alpha\Delta T \tag{2.72}$$

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{2.73}$$

The coefficient of linear expansion, α, is dependent on the sample material and temperature. For example, $\alpha_{aluminium} = 23.1\mu\text{K}^{-1}$ and $\alpha_{steel} = 10.7\mu\text{K}^{-1}$ (carbon steel, 0.7%-1.4% carbon) at

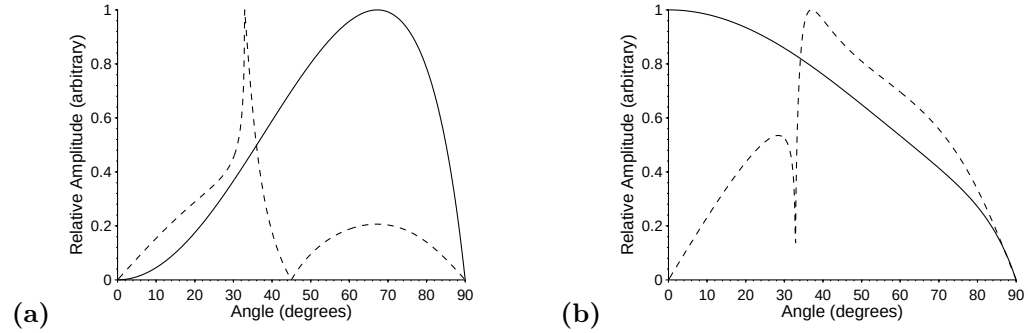


Figure 2.9: Laser directivity for a thermoelastic (a) and ablative (b) source, for compression (solid line) and shear waves (dashed line), with a speed ratio of 1.84 (mild steel) [19]. For the shear wave directivity, the magnitude is taken for the regions in which the directivity is complex [74].

293K [138].

Formulae for the theoretical directivity of a point source, with $\theta = 0$ normal to the surface, and $\theta = \pi/2$ parallel to the surface, are [6, 74]:

$$P_{compression}(\theta) \simeq \frac{2P_0 \sin(2\theta)\sin(\theta)\sqrt{k^2 - \sin^2(\theta)}}{(k^2 - 2\sin^2(\theta))^2 + 4\sin^2(\theta)\sqrt{1-\sin^2(\theta)}\sqrt{k^2-\sin^2(\theta)}} \tag{2.74}$$

$$P_{shear}(\theta) = \frac{P_0 k \sin(4\theta)}{k(1-2\sin^2(\theta))^2 + 4\sin^2(\theta)\sqrt{1-\sin^2(\theta)}\sqrt{1-k^2\sin^2(\theta)}} \tag{2.75}$$

The speed ratio is given by $k = c_p/c_s$. These equations are plotted in figure 2.9 (left).

This is modified for a thermoelastic line source by the following factor, a function of frequency, f, and breadth of the line source, $2b$ (which must be <1mm) [6]:

$$\gamma(\theta) = \frac{\sin(2\pi f b \sin(\theta)/c)}{2\pi f b \sin(\theta)/c} \tag{2.76}$$

Considering some example values, frequency 1MHz, breadth 0.1mm, speed 3000m/s, and angle 80°, the denominator, which is the input for the sine function of the numerator, is sufficiently small that the small angle approximation is approximately valid, and $\gamma(\theta) = 1$ can be used. Indeed, it has been found both theoretically and experimentally [53], that the amplitude directivity of a point source is identical to that of a line source, for the plane perpendicular to the line axis, although the temporal dependence does differ. The equation would suggest that this will break down at higher frequencies, but this is not of concern to this project.

There are models for thermoelastic line sources which better account for thermal diffusion and optical penetration (allowing the laser light to penetrate some way into the sample to form a volumetric source), and include the spatial and temporal distribution of the source [73].

Rayleigh waves are also generated [6]. The peak-to-peak amplitude of both bulk and surface waves is directly proportional to the absorbed laser energy per pulse [6]. Thermoelastic generation from a line source can be used, for example, for estimation of the thickness of thin metal sheet without damaging the sample [59].

2.5.4 Ablative generation

Ablative generation uses the laser to heat the surface to the point at which a small part of the surface ablates (is ionised and expands from the surface in the form of a hot plasma) requiring $I \sim 4 \times 10^{11}\text{W m}^{-2}$ for aluminium and mild steel, achievable with the laser used [6, 7]. The damage can be observed as a mass ejecta crater [132, 139]. The production of vapour introduces a mechanical recoil (far more than from photon reflection [140]) providing elastic energy more directly than the thermoelastic effect, although the relationship is not simple, and the heated area may be many times the focal spot size [74,75,132,139]. The time for which the surface temperature exceeds the evaporation temperature is several orders of magnitude longer than the time for which the laser is incident on the surface, and evaporation takes place over this longer timescale. Plasma is only produced by the metal vapourising followed by further absorption of the incident laser energy. This plasma is highly ionised and expanding rapidly, whereas surface material evaporated later by residual target heat, and by heat transport through the plasma, remains un-ionised and expands at a much slower rate, even if the surface temperature is several thousand degrees above the boiling point. Approximately 90% of the particles ejected per incident laser pulse are neutral atoms which evaporate from the target, but only produce approximately 10% of the stress waves. Approximately 90% of the stress waves are produced by the approximate 10% of the evaporated target material that is formed into a plasma. The fraction of the total ejected mass that appears in ionised form is found by measuring the number of ions in the plasma and comparing it to the total mass lost by the sample [139]. Although the thermoelastic effect is still present, ablative generation will dominate in this regime, and there is a proportionality between laser power density and elastic energy for much of the energy density of interest [74]. Ablative operation has been shown to generate compression, shear and Rayleigh waves in solids [6, 7, 76]. Note that there is a change in polarity of the shear wave between thermoelastic and ablative regimes, and very low shear wave amplitude can be observed at the transition point [54].

The plasma above the surface will not disperse whilst sufficient energy is provided to maintain ablation. The plasma will absorb and reflect laser energy, reducing the energy reaching the surface, and further ionising the plasma causing more absorption of incident energy [6, 132, 139]. This can occur with the high power of the Q-switched laser used, but due to the very short pulse duration of 8ns, the effect is not as pronounced as for a longer pulse. Production of ultrasound requires these factors to be balanced [74]. There is an optimum irradiance for the maximum impulse, below which energy is lost to thermal conduction, and above which energy is lost increasing the temperature and ionisation of the plasma. Thus it is better to use the same energy spread over a larger area to vapourise more of the surface [132].

Formulae for the theoretical directivity, with $\theta = 0$ normal to the surface, and $\theta = \pi/2$ parallel to the surface, are [6, 74]:

$$P_{compression}(\theta) = \frac{2P_0 \cos(\theta)(k^2 - 2\sin^2(\theta))}{(k^2 - 2\sin^2(\theta))^2 + 4\sin^2(\theta)\sqrt{1 - \sin^2(\theta)}\sqrt{k^2 - \sin^2(\theta)}} \tag{2.77}$$

$$P_{shear}(\theta) = \frac{P_0 k \sin(2\theta)\sqrt{1 - k^2\sin^2(\theta)}}{k(1 - 2\sin^2(\theta))^2 + 4\sin^2(\theta)\sqrt{1 - \sin^2(\theta)}\sqrt{1 - k^2\sin^2(\theta)}} \tag{2.78}$$

The speed ratio is given by $k = c_p/c_s$. These equations are plotted in figure 2.9b. Again, wavelength is not included as the source is assumed very small compared with the ultrasonic

wavelength.

At elevated temperatures, less additional energy needs to be delivered by the laser to the surface to reach the ablation threshold. Consequently, the ablative normal forces dominate even more over the in-plane forces that are due to the thermoelastic effect [54].

2.6 Use in development of TOFDI

This chapter has shown how pulsed lasers can be used for non-contact ultrasound generation, and how EMATs can be used for non-contact ultrasound detection (and generation, but they haven't been used in that way within this work). This ability to be physically separated from the sample allows TOFDI, when implemented using such transducers, to be used on rough (and potentially hot) samples, which is a requirement when testing the products of continuous casting before they have been worked. However, it has also been discussed how these transducers can generate and detect multiple wave modes; compression, shear, and Rayleigh waves for example. The different behaviour of these waves, particularly at interfaces, has been described, and as well as being reflected, mode-conversion can occur. This will lead to complicated signals being received, requiring significant signal processing to convert them to defect positions, and prior signal processing stages to removes parts of the signal corresponding to waves that have not been scattered by a defect (chapter 6). The sorts of defects expected and how ultrasound interacts with defects has put the purpose and operation of TOFDI into context.

Chapter 3

Experimental details

This chapter describes how the main TOFDI experiment is set-up, starting with the classic TOFD configuration, adding a pulsed laser ultrasound generator, and multiple EMAT ultrasound detectors. The specifics of these transducers are provided, and how they are configured into a sparse array. Also discussed is the sample, how the scan is controlled, details regarding the digitisation process, how the signal is made complex, how the signal can be filtered, and some sources of noise. These are some of the major points that must be considered when implementing a TOFDI system.

3.1 Experiment set-up

The primary experiment on which this work is based is that of the classic TOFD setup, as depicted in figure 1.1a, modified to perform imaging as shown in figure 3.1.

The sample, which for most of the experiments was a mild steel block (such as that in figure 3.2), was placed on a linear slide. The surface was not polished, but had been milled approximately flat on the top and bottom surfaces.

The linear slide was moved by a stepper motor; an electric motor that can divide its rotation into a large number of steps. Stepper motors are commonly used for driving linear slides or x-y scanning frames within NDT work [10, 20, 26, 30, 31]; because no feedback is necessary, a simple open-loop controller can be used. The motor was connected to a gearhead which in turn was connected to the linear slide by a flexible coupler, specifically a bellows coupling. The 5:1 gear ratio means that the 200 steps per revolution (1.8° per step) in full step drive of the stepper shaft are converted to 1000 steps per revolution (0.36° increment per step) of the linear slide drive shaft. The distance moved by the linear slide per rotation is well known, measured as 1 rotation = (1.60±0.02)mm, or 1mm = (0.63±0.01) rotations, using an average distance per rotation, from (188±1) rotations moving a distance of (300±1)mm (repeated three times) and removing all slack by moving in the measured direction before taking measurements. The result is a distance resolution of 8μm, before gearing. A 5:1 gearing is used, resulting in increased torque, and a distance resolution of 1.6μm per step. The accuracy of the individual steps is far less than the resolution, so such small increments cannot be relied upon. In addition to being controlled by the stepper motor, the linear slide position was tracked by digital callipers with a resolution of 0.01mm, and by a standard low resolution web camera attached via USB to the main control PC (see chapter 4 for details).

It is assumed by the processing used within this project that the emitter-receiver combination is static for a given A-scan. In the laboratory environment, this is actually the case, as the linear

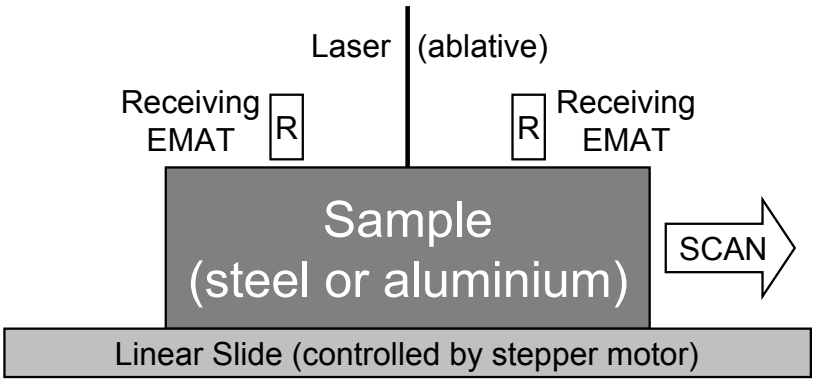


Figure 3.1: The sample is placed on a linear slide, and within the laboratory environment, this can be controlled by a stepper motor. In an industrial environment, this is more likely to be rollers, and the position of the sample must be tracked by other means. The receiving EMATs are placed in the desired configuration around the ablative pulsed laser acoustic generation line source. The laser fires at constant scan increments.

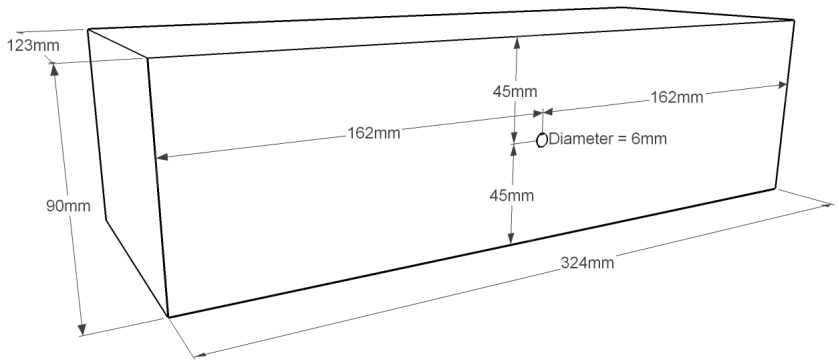


Figure 3.2: A mild steel block of size 324mm × 90mm × 123mm was used as a sample. In the centre of the 324mm × 90mm face is a hole 6mm in diameter, drilled through to the other side. The block is inspected from one of the 324mm × 123mm faces.

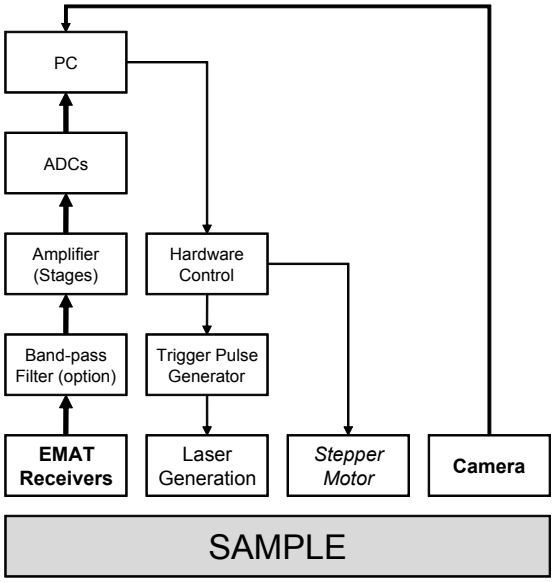


Figure 3.3: The hardware connections for the system. MATLAB will control several ADCs, the sampling frequency of which must satisfy the Nyquist criterion. In this case, the sampling frequency will be over an order of magnitude larger than the highest frequency being recorded. Hardware filters are needed to remove undesirable frequencies. Collections are synchronised with a trigger so that the trigger can be used when the set-up has been moved into the correct position for the next A-Scan (to form a B-Scan).

slide can stop at each scan position. For industrial usage, the assumption is still reasonable, as with a pulse length on the order of 10μs and a collection time on the order of 100μs, at an example speed of 1m/s, the sample would only have moved by 10μm and 100μm respectively. The corresponding worse case Doppler shift (for a 1MHz Rayleigh wave moving along the axis of motion of the receiver/sample at 1m/s) would only be 300Hz.

Receiving EMATs are attached to a rigid aluminium frame which extends over the sample. The receiving EMATs are always of the linearly wound type, although the exact construction varied during the course of the project. Typical lift-off of the EMATs from the surface was $\leq$1mm; variation was due to the sample not having sides that were exactly perpendicular to the plane of the linear slide surface.

In terms of connections, the hardware connections are as figure 3.3. All the hardware is controlled via a single standard PC. For each capture position, first the stepper controller activates the stepper motor to move the linear slide to the desired position. Then the camera takes a photo of the sample for position tracking purposes. Next, the stepper controller fires the laser (typically 4 times), to prepare the surface by removing any light surface contamination, and to allow the ADC to optimise the dynamic range by altering the gain and bias. Then the laser is fired again, to generate ultrasound such that an A-Scan can be recorded, and this is typically repeated 50 times. All the data is stored, converted from an integer ADC level to a voltage, and for each time point, a mean and standard deviation are output (both operations are performed with 64bit precision). This gives an averaged signal with improved resolution and an estimate of the noise at each point on the signal. The cycle repeats for the next capture position. All the data is saved for later analysis.

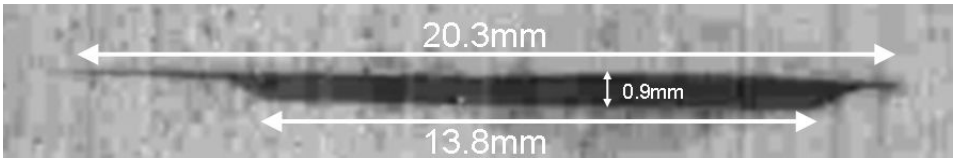


Figure 3.4: A overhead projector transparency was painted black and the laser focused onto it; this burn mark is the result. It is expected that most of the mark is due to ablation of the paint, but some may be due to the secondary action of the plasma formed. The burn mark is 20.3mm long, measuring from its extremities, and 0.9mm wide, measuring from its widest point. The wider part is 13.8mm long.

3.2 Laser ultrasound generation

Ultrasound was generated using a pulsed Nd:YAG (neodymium-doped yttrium aluminium garnet, Nd:$Y_3A_5O_{12}$) laser, the most common type for ultrasound generation [6]. The laser specifications for the Lumonics HY400 state a wavelength of 1064nm, a pulse length of 8ns (Q-switched operation via Pockels cell), and a nominal output of 400mJ. The actual output was measured using a Coherent FieldMaster FM with a LM-P10i pulse detector head; this uses a crystalline pyroelectric material, which acts as a pulse integrator. Pulse integration occurs in the black absorbing coating due to its thermal capacitance, assuming the pulse is much shorter than the thermal time constant of the coating. Setting the detector to 1100nm, the beam was expanded before it was incident on the detector. The energy was measured as (160±5)mJ, only 40% of the nominal output. Despite this, after focusing, the beam still has sufficient energy to act as an ablative ultrasound generation source.

The unfocused beam diameter was measured to be approximately 2mm using simple burn marks upon coloured paper, although the beam shape is clearly not Gaussian. The laser beam was passed through several right-angle prisms before being focused, either to a point on the sample surface using a standard biconvex lens, or to a line on the sample surface, using a plano-concave lens to expand the beam, and a cylindrical lens to focus the beam to a line (figure 3.4). With proper focusing, ablative laser generation was achieved; damage to the sample surface was minimal. Increasing the width of the line source, such that it is less sharp, alters the directivity and the spatial impulse response (SIR). If a point source is used instead of a line source, the beam energy must be attenuated before it passes through the focusing lens, as otherwise the energy density will reach such a high level (even before reaching its focus) that the air will be ionised, and the resultant plasma will absorb a large (but difficult to predict) amount of energy from the beam, as is occurring in figure 3.5. The ultrasonic waves generated through laser ablation are multiple-mode (compression, shear, and Rayleigh waves), and temporally a pulse is approximately 1.5-2.0μs duration.

The first few μs of any A-scans collected using a laser-EMAT system will contain noise due to electromagnetic interference (EMI) from the laser-EMAT RF interaction. EMI due to plasma the laser ablation created, appears later.

3.2.1 Beam splitter and photodetector

When the Nd:YAG laser fires, part of the laser light is diverted via a beam splitter to a photo-detector which provides the trigger signal to the recording device used. The very simple beam splitter used consists of just an uncoated flat glass plate. Placed at 45° to the incident light, most

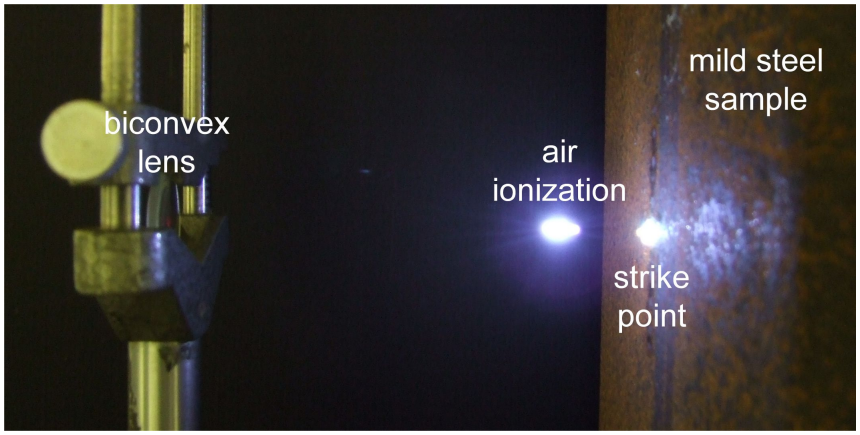


Figure 3.5: A biconvex lens was used to focus a laser beam onto a mild steel sample. However, before the focal point was reached (the strike point on the sample), the energy density of the beam became sufficient to ionise the air. As a result, the ultrasound produced became inconsistent between laser pulses.

of the light is transmitted without change, but a small amount (approximately 4%) is reflected at 90° to the incident angle.

The photodetector used is a BPX65 silicon photodetector, reverse biased, with a rise time of approximately 3.5ns (according to its specifications). Reflectance on the level stated is suitable for triggering from the laser. If the rise time were 4ns, a compressional wave travelling at 6000m/s will only have travelled 24µm between the light reaching the photodetector and the trigger being generated for the recording device.

3.3 EMAT ultrasound reception

The signal from the EMAT passes into a high gain, high impedance, amplifier, with low noise, providing approximately 46dB of gain (following $\mathrm{dB} = 20\log_{10}(A/A_0)$ as is appropriate for amplitude) over the frequency range of interest. The gain was measured empirically using a signal generator creating a 50mV peak-to-peak sine wave (initially at 500kHz, with other frequencies close to this checked), first input into a attenuator, then the amplifier, and finally an oscilloscope. With the attenuator set to zero, and the amplifier bypassed, the oscilloscope measured a 50mV peak-to-peak sine wave, as expected. The operation of the attenuator was then confirmed by checking that the peak-to-peak amplitude measured by the oscilloscope scaled as expected as the attenuator was turned up. By including the amplifier, and setting the attenuator to a series of values between (and including) 60dB and 30dB, the peak-to-peak output from the amplifier was measured as being 46dB higher than the input. Signal distortion appeared negligible, as long as input to the amplifier remained below approximately 3mV, which should include all normal outputs from the EMAT receiver, typically 500µV-5µV.

3.3.1 Electromagnetic interference

The primary points at which noise is introduced are the transducer and the amplifier. The transducer picks up noise from its environment; EMATs are particularly prone to EMI. The amplifier introduces a large amount of noise since any noise added at this stage is amplified with the signal,

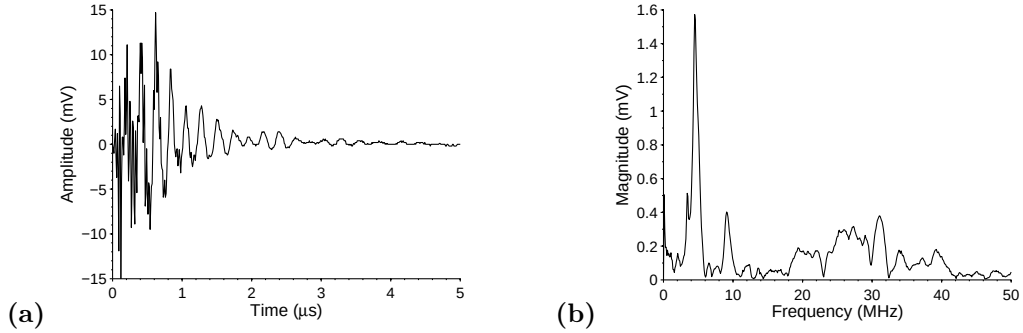


Figure 3.6: An EMAT coil consisting of 30 turns of 0.08mm diameter copper wire around two 12mm cube magnets stacked on top of each other (field direction perpendicular to the surface) detected, reasonably consistently, the signal due to the laser firing (a). This signal arrives too early to be due to an acoustic wave, and doesn't actually need the laser pulse to strike the sample. This would suggest that it is due to RF emissions from the laser electronics. There is frequency content below approximately 40MHz (b). This noise means that the receiver can only detect reliably approximately 4µs after laser pulse emission, corresponding to a distance of approximately 24mm for a surface compression wave on a steel sample. For measurements that can tolerate some interference from the laser, 2µs (12mm) is a less conservative estimate.

whereas noise added afterwards is generally much smaller in comparison to the post-amplification signal. Additionally, the frequency range of the amplifier will overlap with other radio signals (often from local electrical sources), picked up through the mains, intermediate test cables, or earthing points, and these signals may appear in the output signal [75]. Figure 3.6 is an example of RF noise received by an EMAT.

3.3.2 Vibration

Noise can occur due to vibration. An extreme example of this was simulated by hitting the sample directly, and by hitting a nearby connected surface (the table), using a steel hammer with a 2.5cm diameter face. This is not a standard way to generate ultrasound, although it has previously been used to propagate noise (some of which is ultrasonic) [3]. The time and frequency domain responses are shown in figure 3.7. Direct mechanical generation can produce acoustic waves effectively in the 100kHz to 1MHz range (via electromagnetically operated hammers for example) [75], although this has not occurred in this particular test. As a comparison, a laser focused to a point at the same strike point on the sample surface, produced a Rayleigh wave which was detected by the same EMAT, giving a peak-to-peak pulse amplitude of 37mV. The EMAT consisted of 5 turns of 0.08mm copper wire around a 3.5cm diameter cylindrical permanent N52 grade NdFeB magnet with the field direction perpendicular to the surface.

3.4 Array configuration

A B-scan configuration, with the emitter and receivers along a line parallel to the scan, was chosen rather than a D-scan configuration, with the emitter and receivers along a line perpendicular to the scan. Both configurations are shown in a top-down view in figure 3.8. A major reason for this choice was the dimensions of the sample; the sample is much longer than it is wide, and therefore, it is unlikely that more than a single EMAT receiver would fit onto the sample next to

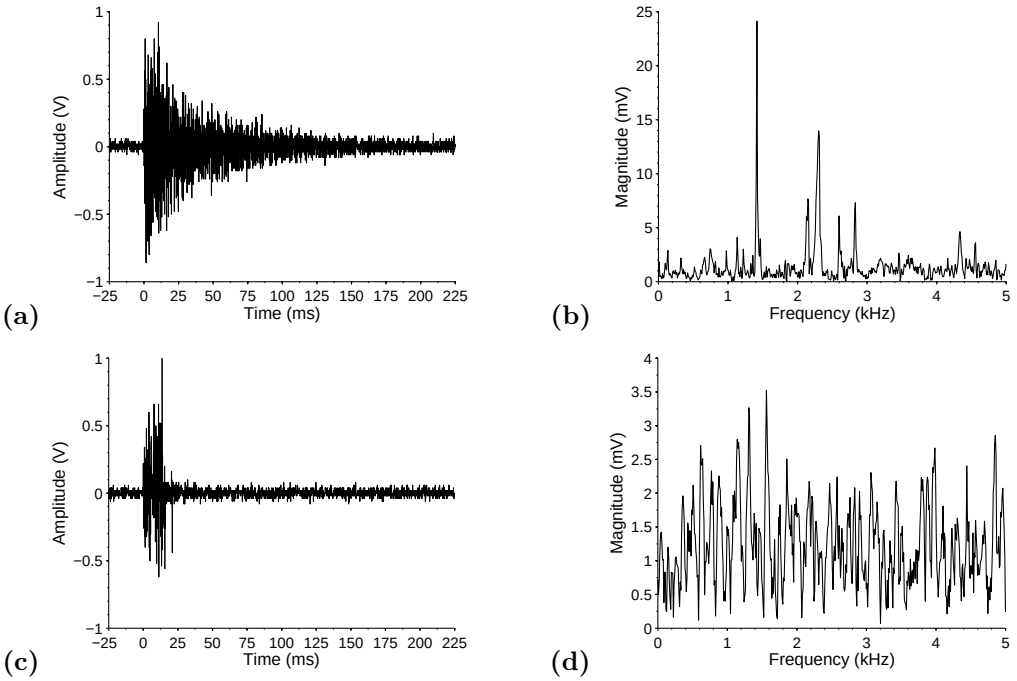


Figure 3.7: A hammer strike produces considerable noise, of a similar magnitude to the signals of interest. When hitting the sample directly (a and b), the point chosen was on the same plane as the EMAT receiver, and was approximately 20cm away, with greater positional accuracy rather difficult due to the manual operation of the hammer. Also tested was striking the surface on which the sample was placed (c and d), at a position of approximately a metre away. This is only intended to give an idea of the noise at the ADC due to a hammer strike, it is not an accurate time or frequency domain representation of the actual hammer strike acoustics, due to the frequency response of the amplifier and EMAT.

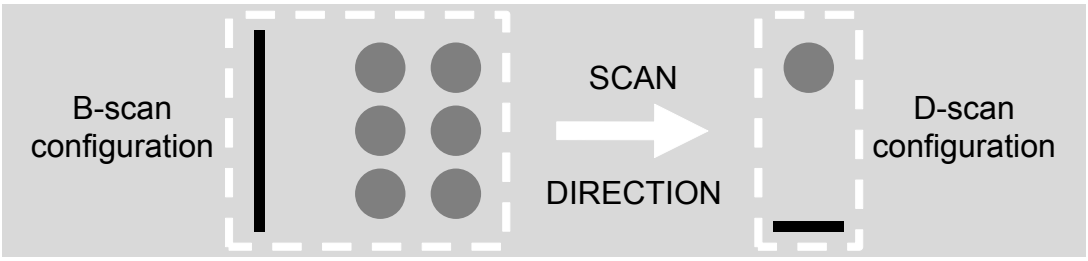


Figure 3.8: A D-scan is a very useful configuration for scanning along, for example, a weld, where the defects are in a known region (somewhere in the weld). However, it is less useful if scanning along an object such as a billet; defects in some regions of the billet are far less likely to be detected. In the diagram, the thick black line represents the laser emitter, and the grey circles are the EMAT receivers. In the B-scan configuration, extra EMATs can easily be added if the directivity of emitter and receiver require it. The D-scan configuration could also be inverted (switching emitter and receiver position), but this would require a second laser source rather than one long one, and firing both lasers simultaneously would cause problems in this case.

the laser emission point if used in the D-scan configuration. This is not a problem in the B-scan configuration, in which multiple rows of receiving EMATs could be fit if necessary, and multiple probe separations are required to cover a range of depths [1]. The directivity of the laser and EMAT would mean that a large section of the edges of the billet in the D-scan case would not be effectively scanned for defects.

3.5 Analog-to-digital conversion

An ADC (or digitiser), of the type used in this project, converts a continuous voltage to a discrete value linearly proportional to the input. The ADC is chosen for its conversion time, resolution (quantisation uncertainty), and linearity. DC coupling is always used as the mean value of the signal is not guaranteed to be zero, and signal droop could occur due to attenuation of low frequencies.

The ADC used for the majority of the work was the GaGe Compuscope 82G waveform digitiser card for PCI bus. Without needlessly providing a full specification, this device could sample up to 1GS/s on two 8bit channels, and provided 250-300MHz of bandwidth. Data could be rapidly transferred to the PC memory. Input impedance was selectable 1MΩ or 50Ω and DC accuracy was ±2% of full scale input. In practice, sampling was performed at 100MHz, and 16384 samples (a collection time of approximately 160μs) were collected after the trigger. Impedance was selected as 1MΩ which allowed input ranges of 0.2V, 0.4V, 1V, 2V, 4V, 10V, and 20V.

Both EMATs and piezoelectric transducers can be used without pre-amplification if the received signal is very strong, but in general, a pre-amplifier is required. This amplifies the signal to a level that the amplifier built into most receivers can amplify further. Where possible, cables have been kept to a minimum length. A long cable can add capacitance to the system, which can decrease the frequency response of the attached transducers [109, 141].

3.5.1 Noise

Noise arises from numerous sources, but for the purposes of this work, the noise is assumed white (thermal/Johnson noise and shot noise can generally be assumed white). White noise has a flat power spectral density (figure 3.9a), although in practice it is bandlimited due to the restrictions of

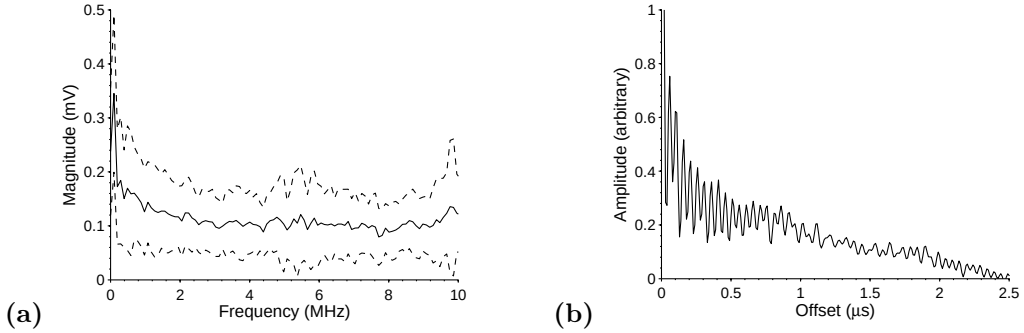


Figure 3.9: A small section of captured data, prior to the arrival of any acoustic waves, was transformed to the frequency domain, and the magnitude was taken. This was repeated for multiple data sets, and the mean and standard deviation of the spectral magnitude at any given frequency was calculated (a). The mean is the solid black line, with the dashed line representing the upper and lower points that are one standard deviation from the mean. The same set of data was autocorrelated (b), and again the mean was taken of the autocorrelation for any given frequency. The steady decrease in the autocorrelation from zero offset would suggest that the signal has some low frequency content that autocorrelates, and indeed a low frequency spike can be seen in (a). Consequently, although the noise is not truly AWGN, it is sufficiently close to be approximated as such.

the receiving system (an oscilloscope has a bandwidth limit for example). White noise is uncorrelated in time (figure 3.9b), resulting in an autocorrelation function which is a Dirac delta function, although again, in practice this cannot be achieved as the noise is bandlimited. The amplitude distribution can take several forms whilst retaining these properties, such as a uniform random distribution (random amplitudes with a uniform distribution), or the Gaussian distribution. Note that a noise source that has a Gaussian amplitude distribution does not have to be white, nor does a white noise source have to be Gaussian, but these two concepts are often paired in what is referred to as additive white Gaussian noise (AWGN), which is a good representation of many real noise sources. In this work, the noise is assumed to be AWGN, although this is not always strictly necessary.

Consider a received signal, $R_i(x)$, consisting of the signal of interest, $S(x)$, and random uncorrelated noise, $N_i(x)$, such that:

$$R_i(x) = S(x) + N_i(x) \tag{3.1}$$

For a signal-to-noise power, $P(x)$, the SNR improvement from taking M averages is given by [1]:

$$P(x) = S^2(x) / < N^2(x) > \tag{3.2}$$

$$SNR_{average} = SNR_{single}\sqrt{M} \tag{3.3}$$

AWGN, regardless of origin (electrical or acoustical), can be reduced using mean averaging (although this does assume an infinite ADC resolution), and this is commonly used within NDT work [10].

3.5.2 ADC resolution

The number of bits, along with the sampling rate, is often one of the ADC specifications that researchers are interested in most. The number of bits defines how many levels a signal is discretised into; 8bit provides $2^8 = 256$ discretised levels, 16bit provides $2^{16} = 65536$ discretised levels. Note that these are maximums, as some oscilloscopes (such as the Tektronix TDS100B series) which are 8bit, only have 255 discretised levels, most likely to provide an even number of levels above and below the zero point.

There is a tendency to assume that a 16bit ADC will always provide a better signal than an 8bit ADC. A signal can be better in several ways, such as a lower error for any given signal level. This can manifest as the accurate representation of both high magnitude and low magnitude signals simultaneously. In the 8bit case, with the 8 bits split into equally sized positive and negative sections, the biggest accurately representable difference in magnitude is 127 levels, whereas for the equivalent 16bit case it is 32767 levels. In both cases, when discussing the largest representable difference, it is assumed there is no noise, whereas in practical cases, the noise level means that the accurately distinguishable difference is less, in that small signals cannot be reliably distinguished from noise. To some extent, the ability to represent accurately both high magnitude and low magnitude signals simultaneously can also be achieved with a logarithmic amplifier (or converter), which converts input signals, V_{in}, to output signals, V_{out}, following the equation $V_{out} = K \ln(V_{in}/V_{ref})$, where V_{ref} and K are constants (note that in practical cases, close to $V_{in} = 0$, such amplifiers cease to behave logarithmically, and most become linear). This results in, for example, a level of 100 above the zero point appearing only twice as large as a level of 10 above the zero point (in both cases the unit is arbitrary), as $\ln(V^2) = 2\ln(V)$, and can be a good solution if the small magnitude signals and the large magnitude signals are temporally separated. An additional way that a lower error for any given signal level can manifest, is that two similar magnitude signals can more easily be distinguished, with one level difference for a 16bit ADC being much smaller than one level difference for an 8bit ADC, if the same overall voltage range is being used. A log amp will not help in distinguishing two similar signals, as $\ln(V^2 + V) \neq 2\ln(V) + \ln(V)$, since clearly a logarithm is not a linear operator. This similar signal level situation will occur if small and large magnitude signals interfere, such that they overlap temporally, and it is necessary to distinguish the section including the small signal.

3.5.2.1 Improving resolution with averaging

This section has been published [142, 143] in a modified form in conference proceedings.

A Monte Carlo model was constructed which can test the expected improvement in the signal level error for an idealised receiver. In order to understand the model, it is necessary to understand how signal resolution can be improved through averaging. If a measurement, x_i, can be repeated N times, where the timing of the signal is known (i.e. it always appears at the same point in time), then repeated measurements can be averaged using an arithmetic mean, $\bar{x}$:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3.4}$$

The population mean is the underlying signal if no noise were present, and the sample mean is an approximation to it. Since the (white) noise is uncorrelated between measurements (i.e. uncorrelated over time, albeit a long time is acceptable in this case due to the gap between consecutive

measurements), and has a mean of zero, as $N \to \infty$, $\bar{x}_{sample} \to \bar{x}_{population}$. If the noise distribution is Gaussian, then the standard deviation, σ, of multiple measurements, will represent the standard deviation of the noise distribution.

In the case of an ADC, the values x_i do not form a continuous distribution, and instead are put into discrete levels defined by the ADC parameters. This introduces a subtlety into the usual relationship for the standard deviation of the sample mean, the usual relationship being:

$$\sigma_{\bar{x}} = \sigma/\sqrt{N} \tag{3.5}$$

The error observed will be a function of where the true signal level is within the discretised level, the size of the discretised level, the noise distribution, and the number of averages, N. For example, if the true signal level is right in the middle of a discretised level, and the size of the discretised level is far larger than the noise in the signal (10σ for example), then the unaveraged signal will be equivalent to the averaged signal. Although it is probable that an analytic model for this additional constraint could be constructed, a Monte Carlo model of the situation can be constructed in very little time, and for that reason, was considered preferable.

The model assesses the error expected on a signal at any given noise level, by simulating AWGN inserted onto a DC signal that is measured multiple times and averaged (arithmetic mean averaging) in subsequent processing. A single discretisation bin of the ADC is chosen in the centre of the discretisation range. The top bin is centred at 0.5V and the bottom bin is centred at -0.5V. Therefore, any given bin covers a range of 3.92mV for 8bit (256 levels) and 0.0153mV for 16bit (65536 levels). The actual choice of discretisation bin is irrelevant as long as the noise does not push the signal above the maximum bin or below the minimum bin, as doing so causes a non-linear response. A series of DC levels are chosen within the selected discretisation bin, ranging from exactly on the input level the bin represents, to the edge of the range the bin represents. Where within the bin the correct value sits, changes the output of the model, as if the input is initially exactly correct, the error can only go up once the noise exceeds the bin range. If the the input is initially at the edge of the range, the initial error is equal to half the size of the bin, but the smallest amount of random noise causes the ADC to switch between the two bins at that edge, and the error quickly decreases before rising again. Most interesting is a point between these extremes, which initially has some error, that slowly decreases as the noise allows the ADC to jump between two discretisation levels, before reaching a minimum error at a threshold noise and then increasing again as the noise increases.

Both 8bit and 16bit ADCs are simulated, for single-shot operation and averaged operation, over 50 captures; the output (figure 3.10) is the difference between the level the ADC outputs and the actual DC input after averaging the ADC outputs over the number of captures (1 and 50 respectively). The experiment is repeated multiple times (averaging the magnitude of the error), until statistical fluctuation due to the random nature of the noise is minimised.

The results for when the DC input is a quarter-way through the discretisation range are shown in figure 3.11. It is assumed that noise never pushes the ADC outside of its maximum range, as otherwise a linear response cannot be maintained. Some statistical fluctuations due to the method of modelling mean that the lines are not perfectly smooth, but this should not detract from the conclusions drawn.

For this input range and taking 50 averages, there is no significant improvement in accuracy using 16bit over 8bit for AWGN with a standard deviation greater than 2mV. However, for single

Figure 3.10: The model takes an input level that sits within one discretisation bin as a DC level and adds a user defined amount of AWGN. The combined signal is digitised in a simulated (ideal) ADC with either 8bit or 16bit resolution. The value output is compared to the exact value input (considering a value with 64bit resolution as effectively exact for this purpose) by simply taking the difference, which is reported as an error. If averaging is being used, the average of multiple outputs from the simulated ADC are averaged with 64bit resolution before the difference is taken and the error reported. This process is itself averaged over multiple iterations to smooth out the variation due to the random nature of the AWGN added.

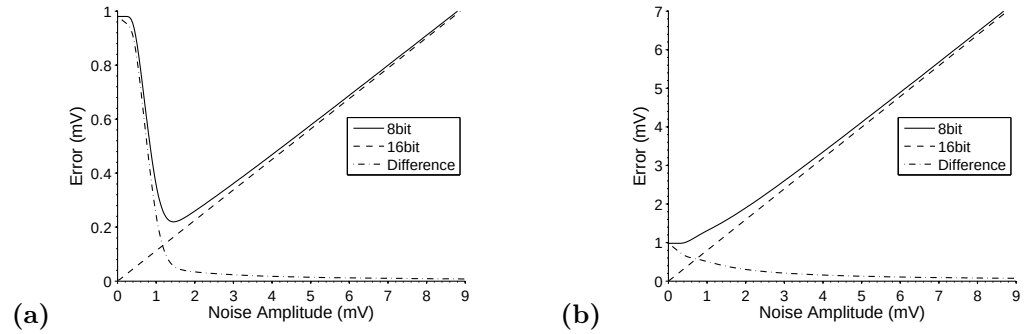


Figure 3.11: Averaging improves the error and resolution. An input value a quarter of the way through a discretisation level is chosen as representative of the worst case averaging improvement in the presence of AWGN. There is a minimal difference (dashed-dotted line) between 16bit (dashed line) and 8bit (solid line) above the threshold noise of approximately 2mV when averaging 50 collections (a), but 16bit is superior in single shot mode (b). Switching to starting centred exactly on a discretisation point, the 16bit ADC never shows any advantage for 50 averages or single shot, but that is a best case and hence not representative. Bin size is 3.92mV for 8bit and 0.015mV for 16bit.

shot operation, there is still a difference of 0.3mV at an AWGN standard deviation of 2mV, and the difference is approximately constant as the noise increases above this. Therefore, within the laboratory environment, no real advantage can be gleaned from a 16bit ADC over an 8bit ADC, as averaging can be used. In industrial use, single shot may be necessary, and this would potentially require the use of a 16bit ADC. If the input signal was ideal and therefore had no noise, averaging would confer no benefit. It is the random noise that allows the improvement via averaging to occur, and if the noise level were actually below the noise level that produced the minimum error for any given number of averages and signal level, then there would be a benefit from artificially adding AWGN to the signal, most likely after amplification but of course prior to digitisation.

For an input range of 400mV, the noise was experimentally tested to be 1.8mV (to one decimal place). On the previous charts, this scales to 4.6mV (to one decimal place). This is easily above the threshold noise, and hence no change in ADC resolution is required if the data is averaged over 50 captures. Data was captured during the preparation laser shots, and this was used to optimise the dynamic range of the ADC by altering the gain and bias such that the signal covered as large an input range as possible, hence using as much of the available resolution as possible.

3.5.3 Conclusions

Noise arises from numerous sources and comes in different forms, but can be modelled fairly accurately in this case by assuming AWGN. Resolution and accuracy (due to a reduction in noise) are improved by averaging multiple measurements of the same time point (relative to the zero time point). Due to the high noise level present in the system, and the averaging employed, there is currently no advantage to a 16bit receiver over an 8bit receiver (all other things being equal). The standard deviation can be recorded in addition to the mean, to make an estimate of the AWGN for each channel; grain-noise will be consistent between captures and hence will not be included in the standard deviation.

3.6 Analytic signal

There are several common ways to get a complex signal, also known as an analytic signal, from a real-valued signal. A way commonly used in NDT [10, 42] is the Hilbert transform (shown here in convolution form):

$$\mathcal{H}\{f(t)\} = (h * f)(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{T} f(t-T)\, dT \tag{3.6}$$

$$h(t) = \frac{1}{\pi t} \tag{3.7}$$

A convolution can be performed by Fourier transforming the signal, multiplying in the frequency domain, and then inverse Fourier transforming. The Fourier transform (FT) is defined by:

$$\mathcal{F}\{f(t)\} = F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt \tag{3.8}$$

The relationship between the FT and convolution is:

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\} \tag{3.9}$$

The inverse Fourier transform (IFT) is:

$$\mathcal{F}^{-1}\{F(\omega)\} = f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega \tag{3.10}$$

The Hilbert transform can then be performed by:

$$\mathcal{H}\{f(t)\} = \mathcal{F}^{-1}\{\mathcal{F}\{h(t)\}\cdot\mathcal{F}\{f(t)\}\} = \mathcal{F}^{-1}\{H(\omega)\cdot F(\omega)\} \tag{3.11}$$

$$H(\omega) = \begin{cases} j & \text{if } \omega < 0 \\ 0 & \text{if } \omega{=}0 \\ -j & \text{if } \omega{>}0 \end{cases} \tag{3.12}$$

This is phase shifting the signal by $\pi/2$. Consequently, the full complex (analytic) signal is then:

$$\mathbf{f}(t) = f(t) + j\mathcal{H}\{f(t)\} \tag{3.13}$$

It must be considered that the Hilbert transform is of infinite support (it is non-zero at every point), and is therefore non-causal as the impulse response exists at negative time, before the impulse arrived. The Hilbert transform is often approximated by a finite support (truncated) filter, time-delayed (such that its impulse response starts at zero time) to make it causal.

A common alternative to performing the Hilbert transform as described, in order to get the analytic signal, is to realise that any real signal can be represented as two complex signals; one with a positive frequency and one with a negative frequency. The positive frequency complex signal is the desired complex signal. This can be obtained by Fourier transforming the real signal, setting all the negative frequencies to zero (hence removing the negative frequency), and then multiplying the positive frequencies by two (since the energy is no longer shared by the positive and negative frequencies). The DC component should neither be set to zero or multiplied by two as it is uniquely represented (the Nyquist component should be similarly treated in practical cases). Inverse Fourier transforming the result then provides the analytic signal. This is simple to describe mathematically. In the frequency domain, a complex signal is given by:

$$S_{complex}(\omega) = \int_{-\infty}^{\infty} Ae^{j(\omega_0 t+\phi)}e^{-j\omega t}\,dt = Ae^{j\phi}\int_{-\infty}^{\infty} e^{-jt(\omega-\omega_0)}\,dt = Ae^{j\phi}\delta(\omega-\omega_0) \tag{3.14}$$

The equivalent real signal in the frequency domain can be described by:

$$S_{real}(\omega) = \frac{1}{2}\left(S_{complex}(\omega) + S^*_{complex}(-\omega)\right) \tag{3.15}$$

To handle the $\omega_0 = 0$ case:

$$S_{complex}(\omega) = \int_{-\infty}^{\infty} Ae^{j\phi}e^{-j\omega t}\,dt = Ae^{j\phi}\int_{-\infty}^{\infty} e^{-j\omega t}\,dt = Ae^{j\phi}\delta(\omega) \tag{3.16}$$

$$S_{complex}(\omega) = A(\cos(\phi) + j\sin(\phi))\delta(\omega) \tag{3.17}$$

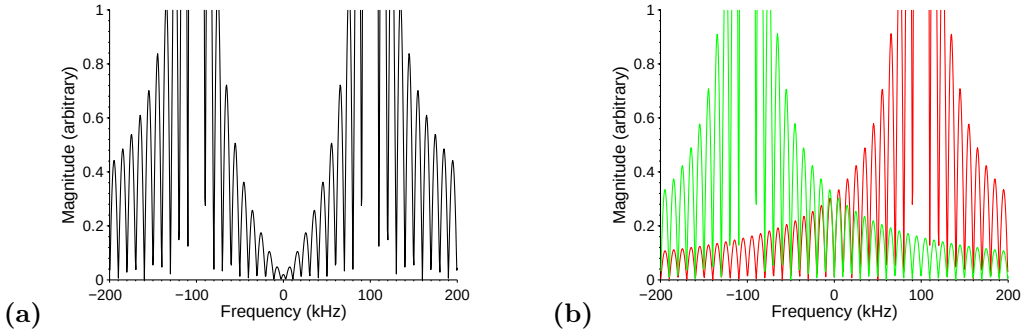


Figure 3.12: The positive and negative sides of the FT can interfere with each other. Taking the FT of 10 cycles of a 100kHz real signal (a), it is clear that the output is not symmetric about 100kHz, although of course it is symmetric about 0kHz. Examining the individual FTs of 10 cycles of ±100kHz complex signals (b), these are symmetric about 100kHz for the positive frequency case, and -100kHz for the negative frequency case.

$$S_{real}(\omega) = \int_{-\infty}^{\infty} A\cos(\phi)e^{-j\omega t}\,dt = A\cos(\phi)\int_{-\infty}^{\infty} e^{-j\omega t}\,dt = A\cos(\phi)\delta(\omega) \tag{3.18}$$

Of course, in practice $A\cos(\phi)$ would only be seen as a single value and interpreted as $A' = A\cos(\phi)$, so effectively $\phi = 0$, and this means that the real and complex transforms are equivalent in this case:

$$S_{complex}(\omega) = S_{real}(\omega) = A\delta(\omega) \tag{3.19}$$

There are some additional implementation issues for the practical case of using a fast Fourier transform (FFT) [144]. The different treatment of the DC term is predicted by the continuous-time infinite-duration form, although it must be derived as a special case. However, the treatment of the Nyquist frequency is not predicted by the continuous-time infinite-duration form (which is required for a correct implementation [144]), but this is only important for transforms of length that are divisible by two as otherwise the Nyquist frequency is not evaluated. Note that FFT algorithms are available that work in $O(N \log N)$ for all N (the number of samples being transformed) and hence odd numbered transforms do occur. For this work, data is usually padded to a power of two if it is not already that size, as such sizes are usually faster, and the Hilbert transform has been implemented taking into account the issues discussed [144]. MATLAB implements the Hilbert transform this way.

Unfortunately, the result of the FFT implemented Hilbert transform will not be perfect. For a real signal with an arbitrary frequency, the positive and negative sides of the FT will interfere with each other, and this can be noticeable at the lower frequencies (figure 3.12). Consequently, the FT of a real signal is not simply a scaled (Hermitian) symmetric version of the positive frequency side of the equivalent positive frequency complex signal FT, which is what has been assumed. In addition, this filter is non-causal (like the Hilbert transform itself) and edge effects can be observed for signals that have not been windowed. However, for must purposes, this filter is sufficient. Obviously it is best suited to data that has been collected in its entirety, rather than data that is being streamed (which would then have to be reformed into a complete signal, potentially with edge effects still present).

Once the complex signal, $s(t)$, has been obtained, the envelope, $A(t)$, can be found by taking

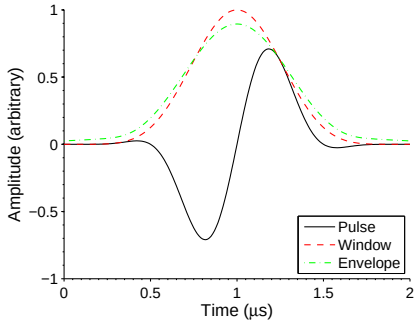


Figure 3.13: The pulse was formed by the multiplication of a sine wave and the window shown. The envelope was generated by taking the complex magnitude of the analytic signal.

the magnitude of the complex signal:

$$A(t) = |s(t)| = \sqrt{s(t)s^*(t)} \tag{3.20}$$

There are other envelope detection methods, but the one described here is fast and sufficiently accurate. In figure 3.13 the envelope generated can be compared to the actual sine wave window, and it is a relatively close match. Also note that now the signal is complex, the sampling rate can be halved whilst retaining all the signal information; the total amount of data remains constant, but previously it was all real, whereas now half is real and half is imaginary.

3.7 Basic filtering

Filtering is used within this project to remove noise. Time-gating is applied to remove any waves that arrive after the last possible single scattering event (such as multiple scatters which this project does not consider), and any signals detected before the fastest wave could have travelled directly from emitter to receiver. The entire capture is also time-shifted to allow the assumption that the centre of the pulse originated at the zero time point. Each A-scan has its median subtracted, to remove DC offsets, assuming the DC offset does not drift significantly during any single A-scan. The median assesses the zero level quite well, as it is only very weakly affected by outliers, such as those caused by the signal from the ultrasound waves, as long as there is a much longer period without significant signal energy present.

3.7.1 Plasma noise

When using a laser-EMAT system (laser generation and EMAT reception) with both generation and reception on the same surface, a wave can be observed due to the EMAT directly detecting the plasma generated when the laser ablates the surface of the sample [139], which a piezoelectric transducer will not detect. The wave can be significantly attenuated by a thin layer of plastic (1mm was tested) placed between the generation point and the EMAT, ensuring no gap is left between the obstruction and the surface.

The wave speed was measured using a B-Scan with the laser generation static and the EMAT moving. By measuring the gradient of the arrival time of the wave, the speed was found to be approximately (373±4)m/s; this is the same order of magnitude as has been measured in other

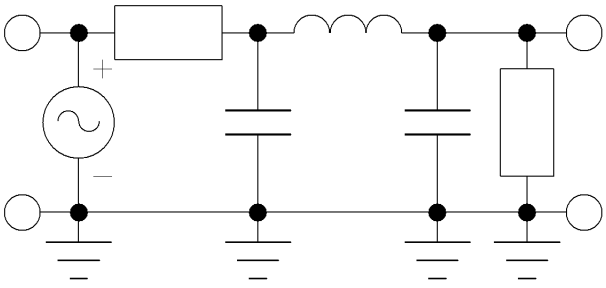


Figure 3.14: Passive analogue low-pass 3-section filter using the Cauer topology.

studies [133], when the plasma has travelled a greater distance, and represents an average speed (the instantaneous speed slowing with distance travelled). However, this method also provided an intercept, as the wave appears before it is expected in absolute terms. The intercept is (19.2±0.4)mm, close to 50μs, which is approximately when the wave reaches the EMAT casing, allowing it to be detected through electrical conduction.

Filtering of the plasma signal is only necessary if ultrasound waves arriving at very late times are required, or if the EMAT receiver and laser ultrasound source are close together, and this is not generally the case. For most cases, the plasma wave can be removed from the signal using simple time-gating. Due to the low frequency of the wave, it is not possible to filter out the wave with a conventional frequency filter, as it would also remove significant parts of the signal of interest.

3.7.2 Low-pass passive analogue filters

The signals received by the EMATs used in this project have frequency domain representations that extend from DC upwards, and hence low-pass filtering is appropriate. Low-pass passive analogue filters ideally contain no resistance, and therefore use mismatch to reflect power outside the pass-band away from the load. Within the pass-band, the low-pass filter matches the load and the source (this approximation is improved for higher section filters). Such filters can be specified by a 3dB point, f_0 [145]. A n-section low pass filter has n = capacitors+inductors. An example of a 3-section filter prototype filter (cut-off frequency $f_0 = 1/2\pi$ radians and source impedance $R = 1\Omega$) is shown in figure 3.14 with normalised values for resistance, $R_{1,2} = 1\Omega$, inductance, $L = 2$H, and capacitance, $C_{1,2} = 1$F. To scale this, resistors become the required load (R'), inductors are multiplied by the required load and divided by the cut-off angular frequency ($L' = LR'/\omega$), and capacitors are divided by the required load and the cut-off angular frequency ($C' = C/R'\omega$). More sections can be included by adding inductors and capacitors to either end (the filter can start with a capacitor and finish with an inductor, or vice versa, or start and finish with a capacitor or inductor) [145].

3.7.3 Band-pass FFT filters

After digitisation, additional finite impulse response (FIR) filtering in software can provide an improved SNR. A band-pass filter is a combination of a low-pass filter and a high-pass filter. If data can be buffered and operated on in segments (these segments are A-scans in this project), it is possible to implement an FIR filter using a standard FFT based technique. The advantage of such a filter is both its mathematical simplicity, and that it can be as complicated as necessary in the frequency domain without any change in the computational complexity. The actual filter

implemented is very simple, and is just a rectangle in the frequency domain that starts at the lowest frequency to pass (typically 0Hz) and ends at the highest frequency to pass (typically 5MHz). For a signal $f(t)$ being filtered between f_1 and f_2:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt \tag{3.21}$$

$$f'(t) = \frac{1}{2\pi}\int_{f_1}^{f_2} F(\omega)e^{j\omega t}\,d\omega \tag{3.22}$$

The filter must symmetrically cover the positive and negative frequencies if the signal is to remain real; many FFT implementations can optionally just examine the positive frequencies and assume the negative frequencies fulfil the required symmetry. Alternatively, this can be combined with taking the analytic signal by multiplying the inverse fast Fourier transform (IFFT) stage by 2 (except for the DC frequency). Care must be taken to not filter significantly any useful frequency components in the signal, as such filtering will adversely affect the time domain representation.

A matched filter was also tried, as an alternative to the band-pass filter; this will detect the signal to which it is matched with the optimum signal-to-noise ratio when the noise can be considered white (the noise must be white if it is to remain uncorrelated after passing through the matched filter) [146]. The measured improvement over the band-pass filter was not sufficient to warrant its use.

3.8 Use in development of TOFDI

This chapter has demonstrated how a TOFDI system can be constructed, and the major considerations when doing so. In particular, the choice of transducers, the properties they must have, and how they are arranged, has been decided for the rest of this work. The B-scan configuration offers advantages over the D-scan configuration when scanning samples that are continuously moved along, as is the case for the prototypical application, and hence has been chosen for this work. Some aspects of this section would not be carried over to the prototypical configuration, such as the linear slide controlled by stepper motors. However, an alternative method for tracking the sample is available, and is discussed in chapter 4. The performance of an 8bit digitiser relative to a 16bit digitiser has been considered, comparing averaged and single shot cases. At the sort of noise levels experienced within this work, there is little advantage to a more expensive 16bit digitiser over an 8bit equivalent, in either averaged or single shot cases. Consequently, the 8bit digitiser is used within this work, and would most likely also be the choice if using TOFDI outside of the laboratory. Finally in this chapter, it has been shown how the signal is filtered, and then made complex using the Hilbert transform described, to improve the imaging process later. It has been shown that the Hilbert transform produces the analytic signal with sufficient accuracy.

Chapter 4

Non-contact position tracking using phase correlation

This chapter has been published [147] in a modified form in a peer-reviewed journal.

This chapter describes a method for tracking a moving sample, that does not require expensive equipment, complicated processing, or any contact with the sample surface. As a primary use of TOFDI will involve a relatively fast moving sample, that may be rough and hot, non-contact tracking may eventually be required.

4.1 Introduction

There are various ways to track the relative position of sample and transducer. The use of a stepper motor to control a linear slide tackles the problem by precisely controlling the sample position, hence tracking the relative position if the transducer is held stationary. A linear or rotary encoder is also commonly employed, sometimes in conjunction with a stepper motor based system [5]. In the same way that some NDT work requires non-contact transducers, these applications and many others could benefit from non-contact position tracking. This chapter considers the use of a basic low-resolution digital camera, in conjunction with the phase correlation technique [148], to remotely track the two dimensional movement and rotation of a sample as it is conveyed.

Phase (as depicted in figure 4.4) is known to be more important than amplitude when considering the intelligibility of an image (figure 4.1). Matched filters that use only the phase information have (for some conditions) been shown to be superior in terms of peak sharpness to those that use phase and amplitude information, although using phase and amplitude typically provides a better SNR in the correlation output, and using just the amplitude information is not viable [149]. However, if only using phase information, the input cannot have a narrow bandwidth if reasonable performance is to be achieved.

Phase correlation can be simply extended to include scale [148], or indeed any other image change that can be converted to a simple shift by means of a coordinate transform [150]. Scale has not been included in this demonstration since checking for changes in scale would not detect changes in height of the sample surface. This is because a change in height is not a change of scale of part of the existing image, but a new part of the image appearing as the sample moves. Therefore, knowing the scale change is not useful when the transducer and sample base are at fixed

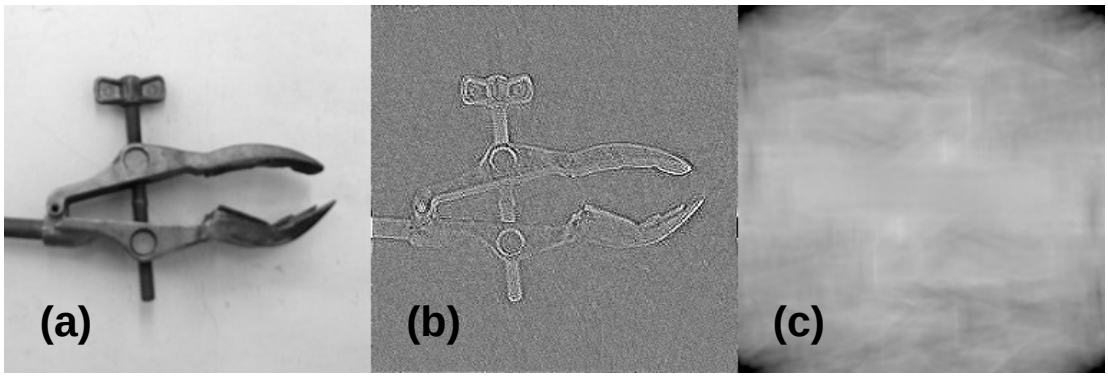


Figure 4.1: Three versions of the same image of a retort clamp. From left, with phase and amplitude (a), with only phase (b), and with only amplitude (c). Due to (b) and (c) having a few very high magnitudes, the greyscale axis was rescaled for these images to improve clarity. Most of the image information appears to be contained in the phase.

heights relative to each other, as was the case for the experimental testing done here. Knowing the rotation is useful, hence it is included in this demonstration. When examining in the frequency domain the cross-correlation of an image with its rotated form, the cross-correlation is only changed from the auto-correlation by a linear phase factor and a rotation equivalent to the angle between the two forms of the image [151, 152].

There are alternatives to phase correlation, such as normalised cross-correlation [153, 154], but phase correlation works well, and can handle both the translation and rotation that could occur, whilst being simple to implement. Such techniques are part of the broader field of image registration, and phase correlation, as used here, is considered an area based method of multitemporal analysis for motion tracking, and is known to be robust to noise [155].

4.2 Theory

The discrete Fourier transform (DFT) is the basis for the FFT, and for a one-dimensional discrete signal, $p(x)$, of size X, the DFT, $P(k)$, for any specific frequency component, k, is given by:

$$P(k) = \sum_{x=0}^{X-1} p(x) e^{-j2\pi kx/X} \tag{4.1}$$

For a two dimensional image, $p(x, y)$, of size (X, Y), the DFT, $P(k, l)$ is obtained for any specific spatial frequency component (k for components along the first dimension and l for components along the second dimension), by taking the DFT of the one-dimensional case:

$$P(k, l) = \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} p(x, y) e^{-j2\pi kx/X} e^{-j2\pi ly/Y} \tag{4.2}$$

Assuming maximum spatial frequency components ($k = K - 1, l = L - 1$), the two-dimensional inverse discrete Fourier transform (IDFT) is then given by:

$$p(x, y) = (1/KL) \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} P(k, l) e^{j2\pi kx/K} e^{j2\pi ly/L} \tag{4.3}$$

Note that $0 \leq x \leq X-1$, $0 \leq y \leq Y-1$, and $0 \leq k \leq K-1$, $0 \leq l \leq L-1$. Usually $X = K$ and $Y = L$, but if zero padding is used, this is not the case, in which case $X \leq K$ and $Y \leq L$. The remainder of this section assumes no zero padding is used unless explicitly stated.

Consider two input images, $p_1(x, y)$ and $p_2(x, y)$, with the latter being a rotated (anticlockwise by θ_0) and shifted (by (x_0, y_0)) version of the former (reference) image:

$$p_2(x, y) = p_1(x\cos\theta_0 - y\sin\theta_0 - x_0, x\sin\theta_0 + y\cos\theta_0 - y_0) \tag{4.4}$$

Re-stating the formulae for rotation and translation invariant phase correlation [148] into the standard form of the DFT implemented cross-correlation [156], and assuming the DFT of $p_1(x, y)$ is $P_1(k, l)$:

$$P_2(k, l) = e^{-j2\pi(kx_0/X + ly_0/Y)} P_1(k\cos\theta_0 - l\sin\theta_0, k\sin\theta_0 + l\cos\theta_0) \tag{4.5}$$

This is obtained using the Fourier rotation theorem and the Fourier shift theorem. If the magnitude of $P_1(k, l)$ is $M_1(k, l)$ then (in Cartesian and polar coordinates respectively):

$$M_2(k, l) = M_1(k\cos\theta_0 - l\sin\theta_0, k\sin\theta_0 + l\cos\theta_0) \tag{4.6}$$

$$M_2(\rho, \theta) = M_1(\rho, \theta - \theta_0) \tag{4.7}$$

If the rotation can be ascertained and removed (by taking the DFT of the magnitudes and using the phase correlation technique described below to identify θ_0), then the shift can be obtained using phase correlation:

$$P_1^*(k, l)P_2(k, l)/|P_1^*(k, l)P_2(k, l)| = e^{-j2\pi(kx_0/X + ly_0/Y)} \tag{4.8}$$

Taking the IDFT to get the phase correlation in the spatial domain, $C(x, y)$:

$$C(x, y) = (1/XY) \sum_{l=0}^{Y-1} \sum_{k=0}^{X-1} e^{-j2\pi(kx_0/X + ly_0/Y)} e^{j2\pi kx/X} e^{j2\pi ly/Y} \tag{4.9}$$

$$C(x, y) = (1/XY) \sum_{l=0}^{Y-1} \sum_{k=0}^{X-1} e^{j2\pi k(x-x_0)/X} e^{j2\pi l(y-y_0)/Y} \tag{4.10}$$

This is a geometric progression, the identity for which is:

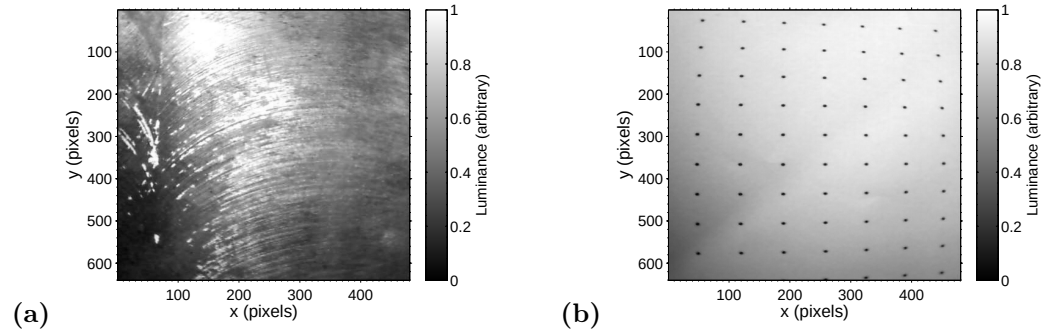
$$\sum_{p=m}^{n} ar^p = a(r^m - r^{n+1})/(1 - r) \tag{4.11}$$

This gives an answer which requires application of l'Hôpital's rule to reach the final result:

$$C(x, y) = \frac{1}{XY} \cdot \frac{1 - e^{j2\pi(x-x_0)}}{1 - e^{j2\pi(x-x_0)/X}} \cdot \frac{1 - e^{j2\pi(y-y_0)}}{1 - e^{j2\pi(y-y_0)/Y}} \tag{4.12}$$

$$C(x, y) = \delta(x - x_0, y - y_0) \tag{4.13}$$

The Fourier shift theorem assumes that $(x_0, y_0) \in \mathbb{Z}$ (this is implicit in using the substitutions $x \to x' = x - x_0$ and $y \to y' = y - y_0$, as otherwise the DFT would have to be performed at



(a) **(b)**

Figure 4.2: The images captured by the camera can have significant perspective distortion at the edges (partially due to camera alignment), although this can be difficult to see in sample images (a). By using a test image of dots at 1cm spacing placed at the same position as the actual sample (b), the perspective distortion can easily be seen, and suitable trimming of the images decided upon to minimise its effect. The trimmed image is figure 4.7a.

fractional sample points), and of course the IDFT input was conjugate symmetric giving a real output.

Shifts that are not whole samples are also possible, and this is the reason for zero padding the IDFT (described in section 4.3.3). If the original signals are interpolated to an extent that the shift is a whole (interpolated) sample, then the previous treatment holds. One practical way to interpolate is by zero padding the FFT of the signal (zero padding the FFT to interpolate is reasonable for a band-limited signal sampled at least at the Nyquist frequency [157]) before using an IFFT to return the interpolated signal. Since both signals require interpolation, it can be done at the correlation stage by zero padding the IFFT after multiplying the signal and the shifted signal, the complex conjugate of one of them being taken prior to multiplication. Consequently, sub-pixel precision correlation can be performed.

4.3 Method

The input image is from a low resolution (640×480, 0.3MP) fixed focus camera returning 8bits per pixel for each of the three colours. The lighting is not uniform across the image but the phase correlation technique is not very sensitive to this issue. For display purposes, the aspect ratio of the images has not been respected. Phase correlation can be optimised to minimise the requirement for the use of the FFT, but this is not described here.

4.3.1 General preparation

Each image loaded (figure 4.2a for example) has the edges removed as the perspective distortion can be quite significant in certain regions. This must be ascertained by the operator by examining test images (such as figure 4.2b) prior to the use of the camera for phase correlation. The image is then converted to greyscale, represented by double precision numbers in the range $-0.5 \leq I \leq 0.5$ (it may be computationally faster to use a different precision, but this is not the most significant concern at this stage). Although optical images used in NDT would have various other image processing operations performed, such as compensation for uneven background illumination [158], this was not found to be necessary here.

4.3.2 Phase correlation for rotational motion

A two-dimensional Hann window is applied to each image to minimise edge effects. This window appears to provide acceptable performance, but few other windows have been tested and there may be superior alternatives for this application. The two-dimensional FFT of the image is taken and only the unique part is retained (see section 4.3.4 for details).

The magnitude of the FFT of each image is then transformed from Cartesian coordinates, (x, y), to polar coordinates, (ρ, θ), ensuring that a suitable radial and angle increment is chosen to retain the relevant image information. Transformations from Cartesian to polar coordinates are performed using bilinear interpolation of the input image. In order to retain all image information, the radial increment should be $\Delta\rho \leq 1$, and the angle increment should be $\Delta\theta \leq \tan^{-1}(y_{max}/x_{max}) - \tan^{-1}([y_{max} - 1]/x_{max})$ assuming $y_{max} \geq x_{max}$, otherwise $\Delta\theta \leq \tan^{-1}(y_{max}/[x_{max} - 1]) - \tan^{-1}(y_{max}/x_{max})$. This assumes that all parts of the image contain useful information, when in practice much of the higher spatial frequency information is not useful as it is either below the noise floor or absent altogether, in which case $\Delta\theta$ can be made significantly larger without detrimental loss of information. The $(0, \theta)$ component can be discarded as it does not vary and hence will not contain information useful for the phase correlation.

The difference between pairs of images transformed will ideally be only along the θ dimension, hence the phase correlation for translational motion purely along one axis (section 4.3.3) can be used to determine the rotation, θ_0.

With the rotation determined, one of the original images (as prepared in section 4.3.1 prior to windowing and FFT operations done in this section) is then rotated using bilinear interpolation, before being checked for translational movement (section 4.3.3). The translational motion check must be in two dimensions since motion along one axis will have been converted into motion along two axes after rotation. The angle output, θ_0, has a 180° ambiguity, which is handled by rotating the image spectrum by θ_0, using phase correlation to obtain the translation, then repeating the process after rotating the image spectrum by $\theta_0 + 180°$, and using the rotation which produces the largest correlation peak [148].

4.3.3 Phase correlation for translational motion

The images are padded along each dimension from an initial length of p to a length of $m = 2p - 1$ to ensure linear correlation (figure 4.3b) later in the process. Without zero padding, the correlation is cyclic (figure 4.3a). If the motion is purely along a single dimension and there is known to be no rotation, then only that dimension requires zero padding. Usually the zero padding is rounded up such that the total size along any direction to be processed by an FFT is a power of two ($m = 2^n$, $n \in \mathbb{Z}^+$) since most FFT implementations are fastest for this size.

The reference image is then windowed for two purposes, split into two windowing operations as described here, but implemented as a single operation. If the maximum expected movement between images is known and is reasonably small relative to the size of the image, a simple rectangular window can be applied that removes parts of the reference image that would not be present in the test image if the maximum movement were to occur. This is combined with a Hann window applied to the remaining image, as otherwise a relatively large peak can occur in the phase correlation output when the edge of the reference image reaches the edge of the test image during phase correlation. The phase correlation operates acceptably without these windowing operations, but performance can be improved under certain conditions by including them.

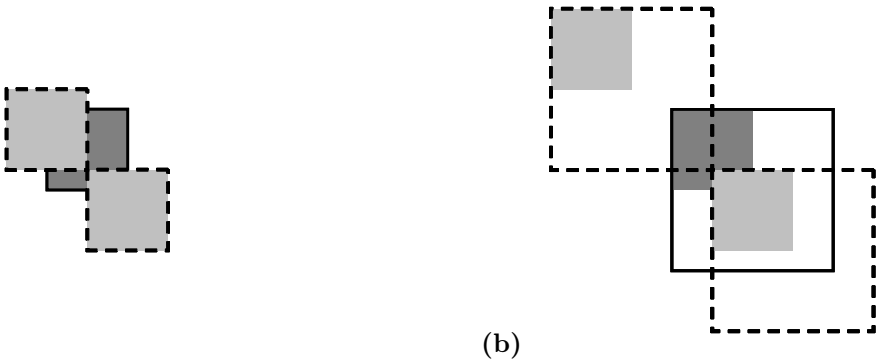


Figure 4.3: The DFT implicitly assumes the image is periodic. Consider two images, represented by boxes that are dark grey (with solid lines) and light grey (with dashed lines) respectively. Without zero padding, using the DFT implemented cross-correlation [156] would result in cyclic correlation (a), such that parts of the correlating image that leave one side would reappear on the other. By padding with zeros, linear correlation (b) is achieved, with parts of the correlating image that leave one side only encountering zeros when they reappear on the other side.

The images are then passed through the FFT, which converts pixel position to spatial frequency; this is usually done along both dimensions, but if the motion is purely along a single dimension and there is known to be no rotation, then only that dimension is operated along (this can result in a considerable computational saving). Only the unique part of the FFT is retained (section 4.3.4) before the complex conjugate is taken of the test image FFT and it is multiplied by the reference image FFT.

Phase correlation whitens the signals by normalisation (figure 4.4), which makes it robust to uniform variations in illumination, offsets in average intensity, and fixed gain errors [159]. However, AWGN extends across all spatial frequencies, and since the use of just the phase information tends to enhance high frequency components, the phase-correlation technique is sensitive to such noise [160]. The reason for this is that a typical image will usually have larger amplitude low frequency components than high frequency components. AWGN has equal components across the frequency spectrum, and hence will have a relatively larger effect upon the smaller high frequency components of the image. Since the phase-correlation technique gives equal weight to all frequency components, the smaller high frequency components will be enhanced to the level of the larger low frequency components, and if noise has affected these components, the result will be a distortion of the correlation output and potentially an error in distinguishing the true correlation peak from false peaks due to noise. Since the image (without noise) is always bandlimited, considering phase components above the bandlimit of the image (the maximum spatial frequency at which the image changes) will amplify the impact of random noise. It is reasonable to ignore frequencies outside the bandwidth of the image when constructing the correlation output [160], although this can unnecessarily widen and shorten the correlation peak if actual image frequencies are filtered away.

Assuming the user has previously examined similar images, components at spatial frequencies which contain signal information of magnitude less than twice as large as the noise at that spatial frequency, can be set to zero to minimise the effect of AWGN on the correlation accuracy. The DC and Nyquist spatial frequencies are always removed as they contain no phase information, since such information at these frequencies manifests as a change in magnitude, not a change in phase. A simple rectangular filter is sufficient since the next step is to remove the magnitude of the cross-correlation by dividing any element with a non-zero magnitude by its magnitude, leaving only the

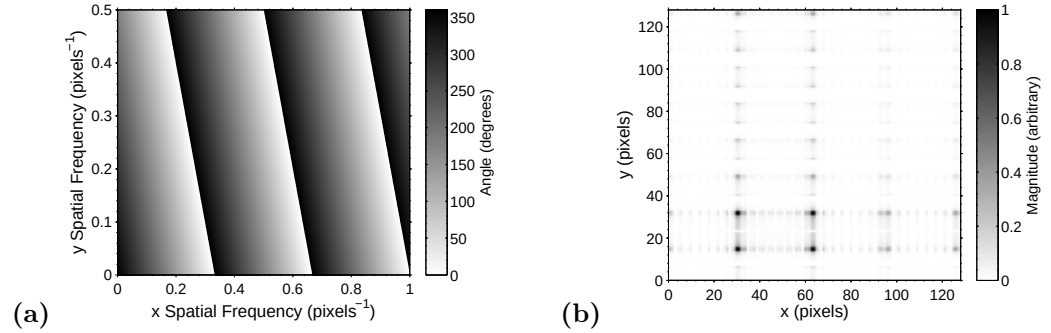


Figure 4.4: The magnitude of the DFT of a single pixel in an image is flat. The phase of the DFT (a) is not flat and is linked to the position of the pixel, here at $(4, 2)$ in a 128×128 image. Consider a rectangle of size 32×16 and bottom left corner at $(31, 16)$ in an 128×128 image. By taking the DFT, normalising such that every frequency component has the same magnitude (whitening the spectrum), and then taking the IDFT, only the phase of the image remains (b). The low frequency components (slowly varying parts of the image) had the largest magnitude, so this process relatively emphasises the higher frequencies (rapidly varying parts of the image). Sharp edges and particularly the corners of the rectangle are emphasised at the expense of the uniform parts of the image. A side effect is that the rectangle gains what are analogous to harmonics.

phase information remaining. An IFFT is then performed to give the spatial phase correlation. The peak can be identified by simply searching for the maximum. It is trivial to convert the peak position to a shift. If a possible range of shifts is known, it is sensible to restrict the search for a maximum peak to this range. The size of the correlation peak can be used to determine the quality of the match.

There are various ways to extend phase correlation for sub-pixel precision [159, 161], but in this case the conventional FFT up-sampling approach is used. For usage on modern computers and small input images, or if only minimal additional resolution is required above pixel resolution, this method is acceptable. All that is required is zero padding the phase correlation in the spatial frequency domain, with each doubling in size, which must be performed along each dimension that was operated on by the FFT, resulting in a doubling of the spatial resolution. Changing the size by doubling it, rather than some other arbitrary multiple, is done to retain the size as a power of two.

For cases where the computational (speed and memory) burden of the conventional FFT up-sampling approach is too great, an alternative method may be necessary, such as the two-step DFT approach [161] using a matrix multiplication implementation of the 2D DFT [162]. There is a significant speed advantage to using the matrix DFT to evaluate the final stage of the phase correlation in an area very local to the expected correlation peak [161]. In the general case, an IFFT without additional zero padding must first be performed to identify the approximate position of the correlation peak, which is still a much faster case than evaluation of the padded IFFT, with significantly less memory required. However, if the shift is known to be limited to a very small range, it could be computationally more efficient to use the matrix DFT to evaluate just this range without first performing an IFFT. In addition, further speed increases can be realised, since the spatial frequency filtering can be incorporated by simply not evaluating terms which would be zero anyway due to the filtering operation. Finally, the DFT matrix approach does not require 2^n samples to run at its fastest (in its standard form), and the outputs do not need to be evenly spaced, removing two of the restrictions of the standard FFT approach, the former restriction

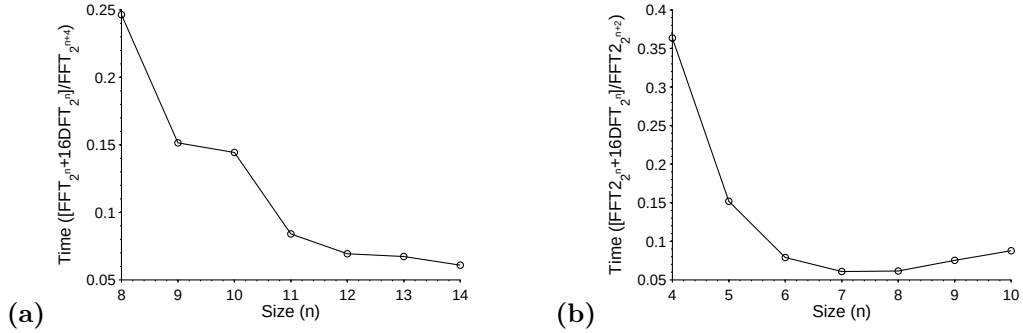


Figure 4.5: Consider a complex 1D FFT output of 2^n samples for which the peak position of the IFFT must be found to $\times 16$ the resolution. This can be interpolated using the standard FFT technique of padding to 2^{n+4} samples before applying the IFFT. Alternatively, the IFFT can be calculated for the 2^n case and then an (inverse) DFT matrix can be applied at the peak to interpolate just that area. The latter technique is relatively much faster (a). Similar speed increases can be seen for the 2D case, even for only $\times 4$ the resolution (b), and the DFT matrix requires significantly less memory. If the original 2D input requires 2^{2n} memory elements, after padding for the standard FFT technique it would require $2^{2(n+4)}$ memory elements, a substantial increase.

being of more significance here.

Figure 4.5 depicts the simulated speed increases for 1D (a) and 2D (b) operation; these simulations assume a sample size of 2^n, so other sample sizes would favour the matrix DFT approach to a greater extent than shown. Values given are for the time taken for a standard size FFT plus a DFT matrix operation, divided by the time taken for a padded size FFT. Values less than unity indicate a speed benefit. The simulations took place on a desktop machine running MATLAB R2007b (32bit edition) on Windows 7 (64bit edition). The PC had an Intel Core2 CPU, model 6420, at 2.13GHz (only using a single core to run the simulations due to limitations of this version of MATLAB), 2GB of RAM (all simulation data fit within this, but it limited the possible size for 2D testing), and no other CPU intensive applications running concurrently. The results are averaged over multiple runs, and deviations were small. The DFT matrix is very simple to implement.

4.3.4 Unique part of the FFT

In the 1D case the unique part of the FFT is the first $m/2$ elements; the $m/2+1$ element is the Nyquist frequency, and is unique, but is discarded during the phase correlation as it contains no phase information. The 2D case is unique for the first $m/2$ elements along the first FFT dimension and all the elements along the second FFT dimension since the second FFT dimension has the FFT operating on complex data rather than the real data of the first dimension. Only retaining unique parts means that computationally expensive operations do not need to be replicated for duplicate data. The IFFT must treat the data as if it were symmetric and of its size prior to discarding information that was not unique, and many FFT implementations have this feature to deal with cases when the data is not exactly conjugate symmetric due to round-off error, a feature that can be abused here to save significant computation.

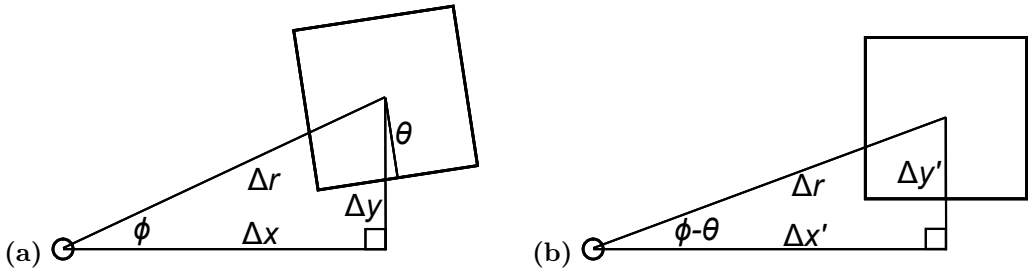


Figure 4.6: The reference image has been rotated anticlockwise by θ and translated by $(\Delta x, \Delta y)$ to form the test image (a), the angle between the image centres is ϕ, and the correct rotation and translation is reported by the phase correlation. If the reference and test images were switched, after the phase correlation rotates the new reference image (b), the angle between the centres is $\phi - \theta$ and the translation is reported as $(\Delta x', \Delta y')$. In practical usage, both cases are equally valid and indistinguishable.

4.3.5 Rotation and translation operation order

When producing test images to check the performance of the phase correlation algorithm, the order by which the reference image was rotated and translated to form the test image can alter the reported translation. Starting with a reference image that is at a rotation of zero degrees and a translation of zero pixels (origin O), the phase correlation technique, as previously described, first adjusts for a rotation by rotating the reference image, and then detects the translation. If the reference image was first rotated and then translated, the reported translation of the test image will match the actual translation performed. If the reference image was translated and then rotated, the reported translation will have the same Pythagorean magnitude, Δr, but the components will differ, since the translation components are also rotated. Consider this second case, and assume a rotation angle of θ and a translation of $(\Delta x, \Delta y)$. Initially the transformed image centre is at an angle of ϕ with respect to the reference image centre (figure 4.6a). After the phase correlation removes the rotation of θ, the transformed image centre is at an angle of $\phi - \theta$ (figure 4.6b). The rotated translation components are then $(\Delta x', \Delta y')$.

$$\Delta r = \sqrt{(\Delta x)^2 + (\Delta y)^2} \tag{4.14}$$

$$\phi = \tan^{-1}(\Delta y / \Delta x) \tag{4.15}$$

$$\Delta x' = \Delta r \cos(\phi - \theta) \tag{4.16}$$

$$\Delta y' = \Delta r \sin(\phi - \theta) \tag{4.17}$$

4.4 Results

Non-contact position tracking using phase correlation has been tested within TOFDI [142, 143, 163]. The results for one-dimensional phase correlation without rotation are for experimental data, whereas the results for two-dimensional phase correlation with rotation are simulated (based on a single experimental input image). The results for operation against noisy images uses the

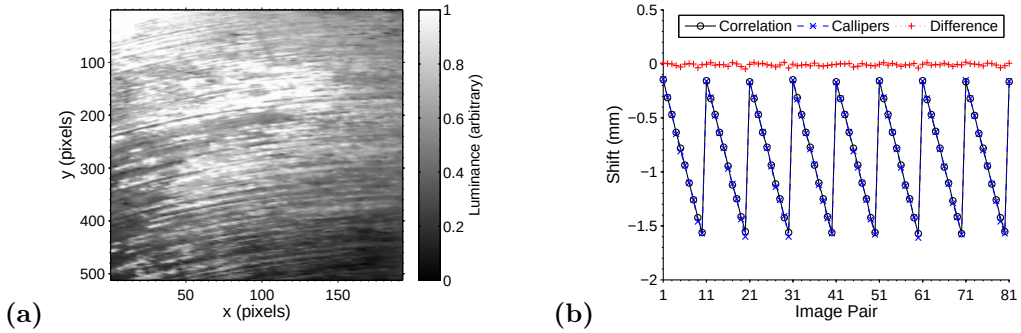


Figure 4.7: At each point at which the position of the sample was measured with callipers, an image of the surface of the steel was taken, moving purely vertically as the image is displayed. This was trimmed to remove the strong perspective distortion at the edges (a), and input into the phase correlation algorithm. The results (b) have black circles representing the shift reported by the phase correlation, blue crosses the measurement by the callipers, and red pluses the difference between the calliper measurements and phase correlation results. The median difference is 0.009mm (0.06 pixels) and the maximum difference is 0.047mm (0.32 pixels).

experimental images used previously, with synthetically added noise.

The images have a resolution of 0.147mm per pixel and are trimmed to remove 30% from each of the left and right edges and 10% from each of the top and bottom edges (based on distortion assessed from figure 4.2b). Note that due to the characteristics of the camera, the image has an aspect ratio far from square. During the phase correlation process, a spatial frequency filter is applied to only allow through components in the range $50m^{-1} \leq s \leq 1250m^{-1}$, based on an analysis of all the input images.

4.4.1 One dimensional phase correlation without rotation

A steel billet was moved along a linear slide by a stepper motor (1.6±0.02mm per 1000 stepper motor steps), producing the repeated step sequence ⌊0.16, 0.32, 0.48, 0.64, 0.80, 0.96, 1.12, 1.28, 1.44, 1.60⌋mm, checked against callipers of accuracy ±0.01mm. An example input image is shown in figure 4.7a. The results for the phase correlation are shown in figure 4.7b. The saw-like shift pattern is due to the step sequence (and hence shifts) increasing by a constant amount until the maximum is reached and the sequence returns to the smallest step.

This is an example of sub-pixel resolution. If the resolution were reduced to the pixel level, the median difference becomes 0.042mm (0.29 pixels) and the maximum difference becomes 0.088mm (0.60 pixels). Usually the answer given by the whole pixel phase correlation will be less than half a pixel from the actual shift, but this still demonstrates that sub-pixel operation does allow for significant precision gains.

4.4.2 Two dimensional phase correlation with rotation

The inputs tested are shown in table 4.1, constructed to separately test rotation in both directions and translation before testing for both transformations together. A pair with no transformation was tested to ensure that a null result was returned when appropriate. The increments used were $\Delta\rho = 1$ and $\Delta\theta = 0.1°$, and sufficient image information was retained in this case, although the limits of these increments were not explored.

Pair	Rotation (°)	Translation (pixels)
0	0.0	(0,0)
1	-15.9	(0,0)
2	15.9	(0,0)
3	0.0	(-33,11)
4	-15.9	(-33,11)

Table 4.1: A single image of the surface of the steel billet was manipulated with an image manipulation program to achieve first the desired rotation and then the translation. A series of image pairs was constructed and input into the phase correlation algorithm, using the original image as the reference each time. In each case, the correct rotation and translation was identified. Switching the reference and test image also produced the correct results, noting the rotation of the translation for pair 4, as described in section 4.3.5.

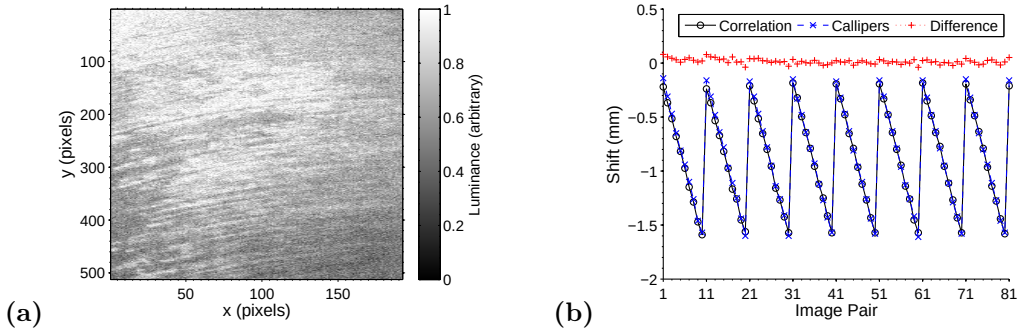


Figure 4.8: This version (a) of figure 4.7a has had AWGN added with a standard deviation of $\sigma = 0.1$ and a mean of $\mu = 0.5$. The image has $\mu \sim 0.5$ over the series of images, giving a signal-to-noise ratio ($SNR = \mu/\sigma$) of $SNR = 5$ on a linear scale. Adding the same amount of noise to every image and phase correlating gives (b). Black circles represent the shift reported by the phase correlation, blue crosses the measurement by the callipers, and red pluses the difference between the calliper measurements and phase correlation results. The median difference is 0.019mm (0.13 pixels) and the maximum difference is 0.081mm (0.55 pixels), which is typical but will vary slightly between runs due to the noise being random.

4.4.3 Operation against noisy images

AWGN was simulated and added to the input images used to obtain the results in section 4.4.1. An example input image is shown in figure 4.8a. The results for the phase correlation are shown in figure 4.8b. Good performance is maintained at this SNR, but drops off more significantly as the SNR becomes even worse. At $\sigma = 0.15$ the median error is typically 0.11mm (0.75 pixels), which is still usable, but at $\sigma = 0.2$ the median error is typically 1mm (8 pixels) and varies considerably between runs, both issues being unacceptable.

4.5 Future work

The next sensible step is to examine the accumulated error in the shift over a large number of image pairs (analogous to accumulated slip in a mechanical system), and if this can be mitigated to some extent by not just comparing consecutive images, but also image pairs representing larger jumps. For example, comparing images 1 and 2, 2 and 3, and 3 and 4, could provide excellent local accuracy, and comparing images 1 and 4 could limit the error due to accumulated errors from comparing consecutive pairs. Changes of scale have not been investigated here, but can be

checked for as well [148], and the accuracy of the technique for scale changes for non-destructive testing applications may be of interest. An example is if the technique was sufficiently accurate to monitor lift-off of an EMAT, moving with the camera in the vertical and horizontal directions, from a top-down view of the sample. Finally, the importance of image resolution should be investigated for a range of surfaces with differing amounts of surface variation and motion blur (which reduces spatial frequency bandwidth). Samples move at significant speed in certain non-destructive testing applications, and it would be of interest to examine how robust the phase correlation technique is to motion blur, or if a very rapid shutter speed, and strong lighting to provide good SNR, would be necessary in these cases.

4.6 Conclusions

There are many applications in NDT which require precise knowledge of the relative position between transducer and surface. This can be done by precisely controlling position changes, using a stepper motor or a servo, but if this is not viable or sufficient, position must be measured using a device such as a linear/rotary encoder, usually requiring contact or very near-contact to operate. An alternative has been demonstrated here, which can be remote from the sample, using a phase correlation tracking technique and a basic camera.

The phase correlation technique is simple to implement, using the FFT for computational speed, and as shown within this work, is able to detect translation in two dimensions and (optionally) rotation, despite the presence of significant noise. It has been extended to sub-pixel precision by conventional FFT up-sampling, and filtering of spatial frequencies outside of those representing real image components reduces the impact of AWGN.

The algorithm has been tested within TOFDI project [142, 143, 163], using only a basic low resolution camera (640×480, fixed focus), tracking a TOFD style transducer configuration [1] as it scanned over a sample surface. Linear sample movement has been determined to sub-pixel precision on experimental data gathered this way, confirmed manually as a median accuracy of 0.01mm of linear movement (0.06 of a pixel) in this case, despite uneven illumination. The performance as noise increased was assessed, and for simulated noise with a standard deviation a fifth of the mean signal level, injected into the experimental data, the median accuracy was 0.02mm of linear movement (0.13 of a pixel) in this case, with some variation in the results due to the random nature of the noise introduced. Rotation was determined to 0.1° on simulated data, but the limits of resolution and noise tolerance were not explored to the same extent as for translational motion.

Future work aims to reduce accumulated error, assess changes of scale, check the importance of image resolution with respect to the spatial frequency bandwidth of the sample surface, and examine the difficulties introduced by motion blur.

4.7 Use in development of TOFDI

This chapter has provided a method for tracking a moving sample, that only requires a basic camera, and relatively simple processing using the phase correlation technique. It is a method that could easily and effectively be used in the prototype application for TOFDI, providing accurate information on the transducer position relative to the sample. The accuracy of 0.01mm is acceptable; a compression wave moving at approximately 6000m/s will take under 2ns to travel that distance. Therefore, in the worse case scenario, the time of flight will be wrong by this much. The

signal amplitude of a 0.5MHz frequency wave (approximately the centre frequency of the generated ultrasound) changes very little over this duration, with a worst case change in amplitude of less than 0.3%. As the worst case change in amplitude is so small, the resultant error in the TOFDI cross-sectional imaging process will negligible. In addition, it is possible that with a better camera and lighting, that greater positional accuracy could be achieved.

Chapter 5

Effect of transducer width on received signal

This chapter describes how the width of a linear coil EMAT changes the frequency response with respect to the receiving angle. Any transducer (piezoelectric, EMAT, or laser) has a finite size for both generation and detection. In certain cases, the size can be considered to be zero (the ideal point emitter/receiver), but in other cases the affect of non-zero size on frequency response must be considered. If a linear coil has more turns, a larger area of the sample will contribute to the ultrasound signal detected, but the larger coil will also lead to increased impedance, and also less sensitivity to higher frequencies. It is important to pick a linear coil with the optimum response for TOFDI, in order to improve the SNR where possible.

5.1 Introduction

There are many factors which result in a change in the frequency response of the EMAT, such as the magnet shape and strength, and the coil configuration (shape and number of turns being important factors). In particular, the latter can easily lead to a change in impedance, which is dependent on frequency, as is inductance [109, 141], which would also change.

A linear coil has been chosen for this TOFDI implementation, because that coil configuration can detect bulk ultrasound waves (compression and shear) over a relatively wide range of angles. In terms of bulk wave EMATs, the coil configurations used are usually either linear, spiral, or meander line. The meander line coil can be used to focus SV waves onto a focal line, which is not appropriate for the TOFDI application. The spiral coil usually has good performance close to (but not at) normal incidence, and is very useful for pulse-echo work, but has much poorer performance further away from normal incidence. In addition, the spiral and meander coil may not effectively detect plane waves that extend over a significant portion of the coil, due to the symmetry of the coil causing the plane wave to cancel itself; if equivalent motion is detected in two opposing parts of the coil, these parts will cancel and the sum will be zero or very small, depending on the exact scenario. A linear coil does not suffer from this limitation, and is therefore the choice for detecting bulk waves in the prototypical TOFDI application. Once a linear coil (such as in figure 2.5) has been selected, the magnetic field should be one that is strong and relatively consistent over the area of the coil, with the field direction dictating which oscillation direction is detected. As it is

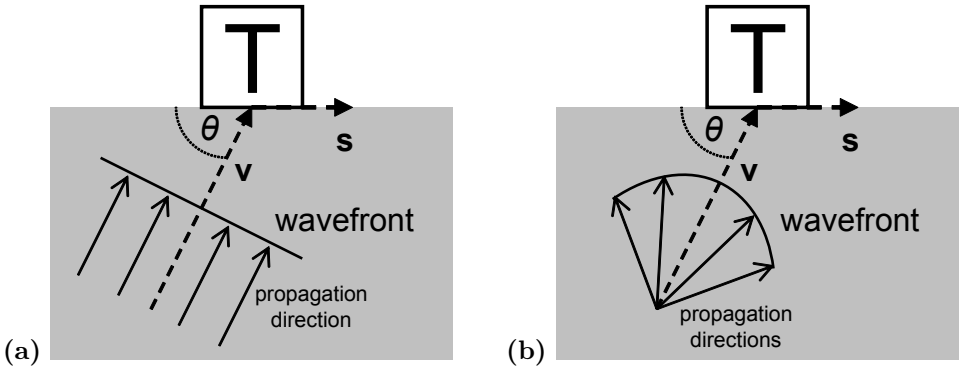


Figure 5.1: The plane wavefront (a) represents the far-field approximation. The circular wavefront (b) represents the near-field approximation. The plane wavefront form is a good approximation in many cases where the deviation in time of flight and amplitude can be considered negligible for the section of the wavefront received by the transducer. The wavefront polarisation is perpendicular to the direction of propagation for shear waves, and parallel to the direction of propagation for compression waves. If the transducer is sensitive to the small variation in polarisation over the wavefront, the circular wavefront form may need to be considered even at distances that might normally be considered the far-field.

far easier to create a strong magnetic field perpendicular to the sample, that direction was chosen for this work (figure 2.7), although there are deviations due to the magnet not having an infinite extent (edge effects).

Having chosen the other key EMAT configuration parameters, coil width must be considered. For a series of different linear coil widths, each corresponding to a different number of turns of wire, this chapter explores the sensitivity of each coil to different signal frequencies with respect to angle of incidence of the ultrasound, and checks that it approximately correlates with a simple model.

5.2 Theory and method

First consider a receiver of finite width, the exact operation of which is not important here. Assume the incoming wave is in the far-field and hence the angle of the wave front, θ (with 0° being parallel to the surface plane, and 90° being perpendicular to the surface plane), with respect to the receiver face, is constant across the face of the receiver. Also assume that the scenario can essentially be considered two-dimensional; an infinitely long transducer with finite width with a wave travelling in any direction perpendicular to the dimension of infinite length, figure 5.1a.

The wave travels past the transducer at a rate dependent on the velocity, $\mathbf{v}$ (the speed is $v = |\mathbf{v}|$), and the direction of travel relative to the transducer size vector, $\mathbf{s}$. This can be considered as an effective transducer width s' (figure 5.1a):

$$s' = \mathbf{s}.\hat{\mathbf{v}} = |\mathbf{s}| \cos\theta \tag{5.1}$$

The finite size of the transducer would act as a filter, in this case a rectangular filter. This leads to the filter (symmetric about $t = 0$):

$$f(t) = \sqcap(t) = \begin{cases} 0 & \text{if } |t| > |s'/2v| \\ 1 & \text{if } |t| \leq |s'/2v| \end{cases} \tag{5.2}$$

The consequence of this is that the wave will be observed before it reaches the centre of the transducer, as of course the edges of the transducer respond first. An ideal signal represented by a Dirac delta function would, after detection, assume the form of the filter (from basic convolution). Note that if the angle of incidence is perpendicular to the face of the transducer, the effective transducer width is zero, assuming that the transducer has no depth (a sufficient approximation at this stage). Therefore, this is the only angle of incidence for which the signal is not made temporally wider.

The Fourier transformation of the filter (with width $x = |\mathbf{s}| \cos\theta / 2v$) provides the frequency response:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt \tag{5.3}$$

$$= \int_{-x}^{x} e^{-j\omega t}\, dt \tag{5.4}$$

$$= \frac{2}{\omega} \sin(\omega x) \tag{5.5}$$

$$F(\omega) = \frac{2}{\omega} \sin\left(\frac{|\mathbf{s}| \cos\theta}{2v}\omega\right) \tag{5.6}$$

Factors such as receiver sensitivity have been neglected. The finite size does not, in this case, alter the phase, and this is what you would usually expect. The equation breaks down if $\theta = \pi/2$, and in this case the filter is a Dirac delta function $\delta(t)$ and there is no change to the received signal frequency spectrum (the ideal case).

$$F(\omega) = \int_{-\infty}^{\infty} \delta(t) e^{-j\omega t}\, dt = 1 \tag{5.7}$$

The point at $\omega = 0$ needs special care when calculating, and can be shown using l'Hôpital's rule to be:

$$F(0) = \left(\frac{|\mathbf{s}| \cos\theta}{v}\right) \cos\left(\frac{|\mathbf{s}| \cos\theta}{2v}\omega\right) = \frac{|\mathbf{s}| \cos\theta}{v} \tag{5.8}$$

Examining equation 5.6, it can be seen that as $\theta \to \pi/2$ the response becomes vanishingly small. This is because the wave front has been modelled as an infinitely narrow ray rather than a broad wavefront. To include a wavefront broad enough to cover the entire transducer face, the rectangular function should maintain a constant area as the angle varies; if the entire face of the transducer is illuminated, the same amplitude can be observed regardless of direction for a DC wave (a convenient theoretical construct in this case). The filter width is $|\mathbf{s}| \cos\theta / v$ and the filter area can be defined as unity, therefore dividing by $|\mathbf{s}| \cos\theta / v$ will keep the area constant. If the expected response for the theoretical DC wave is proportional to the actual width $|\mathbf{s}|$ (with some sensitivity factor R as the constant of proportionality), this should also be included, such that the transducer response equation normalisation factor (P) is:

$$P = R\,|\mathbf{s}|\, v / |\mathbf{s}| \cos\theta = Rv / \cos\theta \tag{5.9}$$

Multiplying this factor into equation 5.6 gives the final transducer response equation (including the zero frequency case which must be obtained using l'Hôpital's rule):

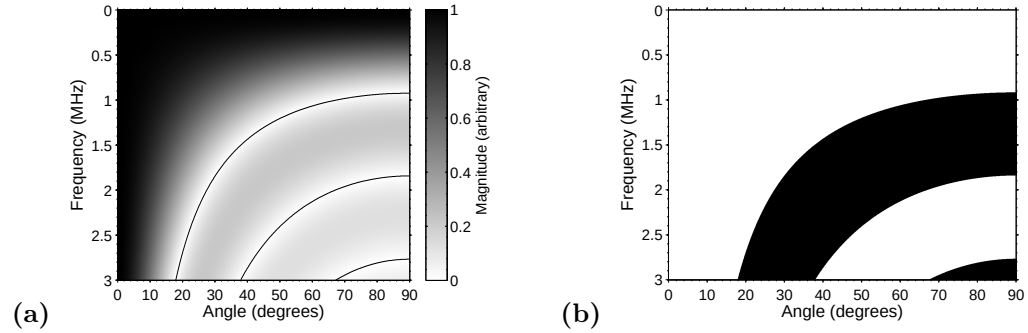


Figure 5.2: The simulated frequency response for a plane wave incident onto a 30 turn EMAT (coil width 3.5mm from table 5.1), placed on a steel sample. The wave speed was taken to be 3226m/s. The magnitude (a) of the frequency response shows three troughs, marked by thin black lines. At each trough, the phase switches by 180°, forming phase reversal regions, marked black in the image showing phase (b). Compared to the signal received when the incident angle is 0°, in these phase reversal regions, not only is the spectral frequency magnitude attenuated as shown in (a), but it is also 180° out of phase.

$$F(\omega,\theta) = \frac{2Rv}{\omega\cos\theta}\sin\left(\frac{|\mathbf{s}|\,\omega\cos\theta}{2v}\right) \tag{5.10}$$

$$F(0,\theta) = F(\omega,\frac{\pi}{2}) = R\,|\mathbf{s}| \tag{5.11}$$

The signal is phase shifted by π if; the transducer is facing the other way, or the wave is travelling in the opposite direction, or $\theta > \pi/2$ (equivalent to the previous two statements), or $\omega = 4\pi v/s\cos\theta$. The final one of these defines the first trough in the frequency response, and the phase is shifted by π until the next trough, when it is phase shifted back. This repeats until the signal is no longer detectable. Note that these are troughs in the frequency response due to the size of the transducer, there may be other troughs due to the underlying frequency content of the signal. This is simply what would be seen for a white-noise-like signal. Examining the frequencies at which troughs occur, and the phase shifts at these frequency points, is the most accessible way to test this theory.

In three dimensions, the situation is similar, and considering the simple case of a circularly symmetric receiver, such as a piezoelectric with a circular face, the filter is then representing the amount of transducer seen as the wave passes by (again assuming the far-field case).

$$f(t) = \begin{cases} 0 & \text{if } |t| > s'/2v \\ \cos\left(\frac{2v}{s'}\pi t\right) & \text{if } |t| \le s'/2v \end{cases} \tag{5.12}$$

A transducer has an effective depth that can often be considered negligible, such as the electromagnetic skin depth when using an EMAT. The two dimensional case is a reasonable approximation for a linear coil EMAT as long as edge effects are ignored and considering that $\mathbf{s}.\hat{\mathbf{v}}$ is a function of (θ,ϕ) if working in spherical polar coordinates. Figure 5.2 shows an example output from this model.

To apply this analysis, by the time the wave front reaches the receiver, it must be uniform to invoke the far-field approximation and to ignore emission edge effects. Incorporating the near-field (figure 5.1b) would require a different filter shape, a shape dependent on the relative positions of

emitter and receiver and their characteristics, such as directivity; as this is more complicated, it is instead handled computationally.

In this case, the emitter is considered to be point-like; multiple emitters can be combined if necessary, but a single emission element seems to well represent a sharp laser ultrasound line source. This is because the simulation is two dimensional, and the laser ultrasound line source is experimentally longer than the EMAT receiver, making the experiment approximate the two dimensional simulation in some cases. Therefore, a single point represents an infinitely long laser ultrasound line source of very narrow width. As the experimental laser line source is much narrower than the EMAT coil width, the approximation holds under some conditions. To achieve a sharp laser ultrasound line source, the laser beam is expanded and then focused into a line using a cylindrical lens, with the line the same orientation as the EMAT coil, and focused onto the sample such that the focal point is as close to the surface as possible, but just past the surface rather than just before.

The receiver is considered as a series of point-like receivers; the equivalent experimental setup is shown in figure 5.3. That figure shows the generated wave propagating into the sample (the dashed lines represent example paths of the emitted wavelets), and being received by the EMAT (split into discrete point-like receivers at the surface within the simulation). The distance and therefore time and geometric attenuation from the emitter point to each receiver point is calculated. Assume an emitter that sits on the semi-circular sample (of radius r) at $E_x = r\sin(\theta)$ and $E_y = r\cos(\theta)$. Each receiver point is at $R_y = 0$ and R_{x_n}, ranging from $R_{x_1} \leq R_{x_n} \leq R_{x_N}$. The centre of the receiver is at $(0, 0)$. The range of angles this covers is then:

$$\tan\left([R_{x_1} - E_x]/E_y\right)^{-1} \leq \phi_n \leq \tan\left([R_{x_N} - E_x]/E_y\right)^{-1} \tag{5.13}$$

The x distance to each point on the receiver for each angle ϕ_n is calculated by $R_{x_n} = E_y \tan\phi_n$. The distance to each point is then given by:

$$d_n = \sqrt{(R_{x_n} - E_x)^2 + E_y^2} \tag{5.14}$$

At this stage, a compression wave, and a receiver that only detects motion that is in-plane, is assumed (the EMAT used for this test is optimised for in-plane motion). Note that compression and shear waves have in-plane and out-of-plane oscillations, as this is a function of the propagation direction (compression waves oscillating parallel to the propagation direction, shear waves oscillating perpendicular to the propagation direction), and so this assumption is stating that the receiver is predominantly only sensitive to the in-plane oscillation of a given wavefront. Therefore, the ratio x_n/d_n informs you of how much of the motion of the compression wave is in-plane (relative to the receiver position, figure 5.3) if the total motion amplitude was unity. In addition, the ratio also informs you if some of the in-plane motion is in a different direction. Consider the $E_x = 0$, $E_y = r$ case; half the incident wave has motion in the $-x$ direction, and half has motion in the $+x$ direction, and hence the sum will cancel perfectly.

For each receiver point, the example signal, either obtained experimentally or simulated, is shifted in time by d_n/c, with c the speed of the compression wave, and the amplitude scaled by x_n/d_n (as well as appropriate scaling for geometric spreading of the signal). The signal for each receiver point is then summed to get the total signal for the receiver for an emitter at that point. This is repeated for a series of emitters at angles in the range $0° \leq \theta < 90°$. With the received signal in the temporal domain calculated, the frequency domain signal is calculated by taking

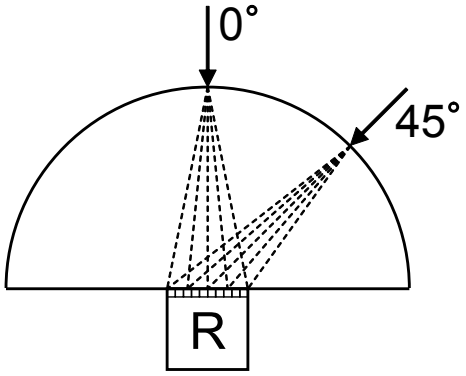


Figure 5.3: This is a top-down view of the experimental set-up used for EMAT directivity measurements. Semi-cylindrical blocks of aluminium and steel were made (radial dimension of 5cm), that were long in length (into the page). An EMAT was placed on the flat side (labelled "R"), with the wires running into the page such that the width of the EMAT coil is parallel to the flat edge of the semi-cylinder that is shown. The laser source is focused onto the curved edge of the semi-cylindrical block. Using a turntable, the sample is rotated in place so that the laser always strikes perpendicular to the curved edge.

the Fourier transform of the temporal domain signal. For display purposes, the FT magnitude is shown, and the magnitude at each angle is scaled such that the maximum magnitude for that angle is equal to one.

5.2.1 Compression and shear wavefront shapes from a line source

It has already been discussed that in the $E_x = 0$, $E_y = r$ case; half the incident wave has motion in the $-x$ direction, and half has motion in the $+x$ direction, and hence the sum will cancel perfectly. This matches what is seen experimentally, but in addition, FEA (PZFlex) was used to examine differences in pulse shapes. The sample modelled was effectively a semi-infinite two-dimensional surface. Within the simulation, the receiver is only a short distance from the emitter, and the waves are not reflected, diffracted, or in any other way altered. Reception points are parallel to the surface, but offset from it into the depth of the sample. Velocities perpendicular and parallel to the surface are considered separately. Simulated (in separate trials) are a piston-like (uni-directional) force applied perpendicular to the surface, and a uni-directional force applied parallel to the surface, both applied to a single element at the surface of the modelled sample, both creating compression and shear waves. Temporally, the signal consists of two complete cycles of a 1MHz sine wave multiplied by a simple window function.

Considering the phase when the applied pressure load is perpendicular to the surface (figure 5.4), both compression and shear waves have the same phase either side of the emitter when observing the velocity perpendicular to the surface, and the wave types have the same phase as each other as well. When observing the velocity parallel to the surface; both the compression and shear waves have a different phase either side of the emitter. For the compression wave, the phase matches the phase seen for the velocity perpendicular to the surface, on the right side of the emitter; the opposite is true for the shear wave. Since the polarisation must change between the right and left of the emitter, the shear wave has no component parallel to the surface at the reception node directly below the surface.

Considering the phase when the applied pressure load is parallel to the surface (figure 5.5), both

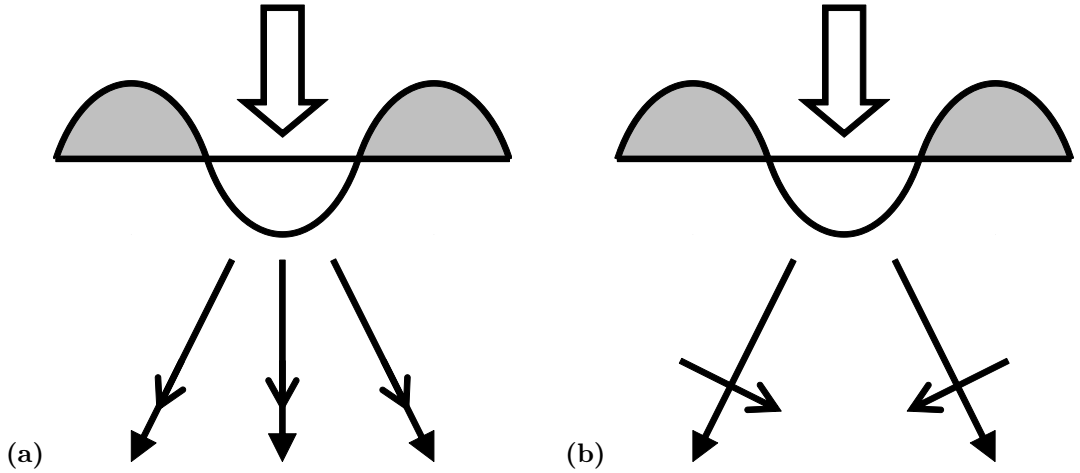


Figure 5.4: Wave polarisations, longitudinal (a), and transverse (b), resulting from a piston-like (uni-directional) force applied perpendicular to the surface. The oscillation velocities perpendicular and parallel to the surface are symmetric along the centre line.

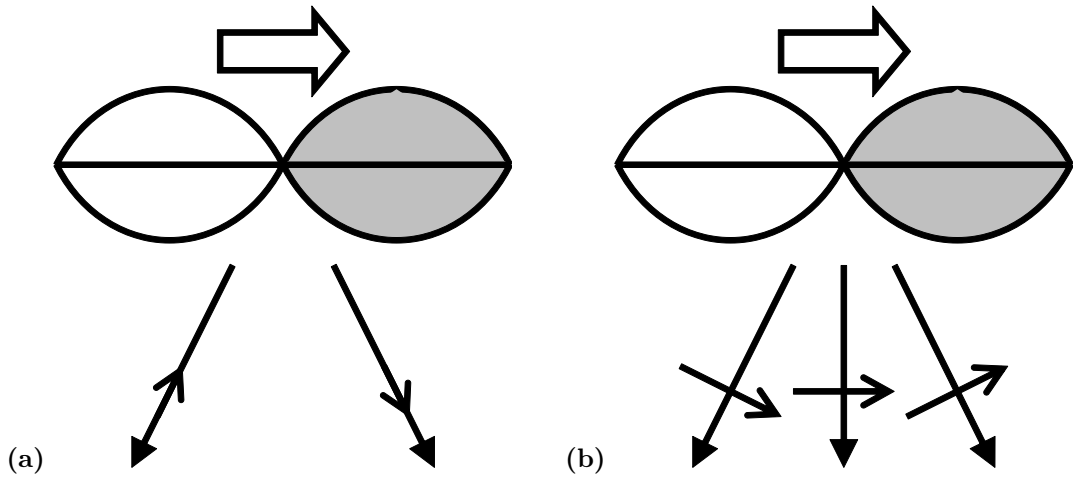


Figure 5.5: Wave polarisations, longitudinal (a), and transverse (b), resulting from a uni-directional force applied parallel to the surface. The oscillation velocities perpendicular and parallel to the surface are antisymmetric along the centre line.

compression and shear waves have the opposite phase either side of the emitter when observing the velocity perpendicular to the surface, and the wave types have the opposite phase to each other as well. When observing the velocity parallel to the surface; both the compression and shear waves have the same phase either side of the emitter. For the compression wave, the phase matches the phase seen for the velocity perpendicular to the surface, on the right side of the emitter; the opposite is true for the shear wave. Since the polarisation must change between the right and left of the emitter, the compression wave has no component parallel to the surface at the reception node directly below the surface. Note that spiral coil EMATs are also known to produce shear waves with zero amplitude along the axis of the generation EMAT [164].

Turns	Width (±0.2mm)	Resistance (±0.1Ω)
10	1.0	3.6
20	2.0	7.0
30	3.5	9.5
40	4.5	12.1
50	6.4	15.5
60	7.0	17.9
70	8.9	22.8

Table 5.1: A series of EMAT coils were constructed for directivity testing by winding insulated copper wire of 0.08mm diameter around a rod of diameter (24±0.1)mm. The number of turns around the rod, the total width of the wire along the rod, and the resistance of the coil were recorded. A layer of tape was used as a base on which the wire adhered, and as a cover over the wires once the coil was wound. The coils were removed from the rod and magnets were placed within the coils to form receiving EMATs.

5.3 EMAT construction

The primary EMATs tested in this section used cuboid NdFeB magnets of grade N48. Each side had a length of (12±0.1)mm. The magnets were nickel-plated (Ni-Cu-Ni) and have a useful maximum working temperature of approximately 350K. For these EMATs, the magnets were stacked two high. Measuring the flux density at the centre of a North/South face on a combined magnet stack, using a Hall probe with an accuracy of ±1mT, gave a reading of (600±10)mT. An individual magnet gave a reading of (540±5)mT, the variance being due to variation between magnet pairs, as several were measured, and the mean and standard deviation taken. Such variation has to be expected, and <2% seems very reasonable. Considering the future construction of EMATs using such magnets, variation of 5% (±30mT) should be tolerable. Realistically, 0.5mm is as close as the Hall probe can measure (the cover for the sensor is likely to be approximately this thickness on its own). It should also be noted that the Hall probe will be measuring the average field over a certain area, and hence will not measure the field exactly at its strongest point (if the magnet is in free space); a point exactly at the centre of a North/South face.

The EMAT coil consisted of 0.08mm diameter insulated copper wire. The wire was wound around a "winding rod" of diameter (24±0.1)mm. The combined exterior length of the stacked magnets was (72±0.2)mm and the circumference of the rod was (75±0.3)mm. The extra 3mm allowed the coil to be put (easily) onto the combined magnet, without placing it under tension. It was ensured that the coil was flat against the side of the magnet being used for detection of ultrasonic waves, and the coil was also flat against the adjacent sides of the magnet. The gap was entirely at the "top" side, furthest from the detection surface. This was also the side at which the ends of the wire were soldered to a BNC connector. The DC resistance of each coil was measured, as was the width of each coil. The width of each coil represents the extent in the dimension perpendicular to the wire winding direction on the detection surface. This information is recorded in table 5.1.

All the parts of the EMAT described were placed inside a brass case for easier handling, improved durability, and electrical screening.

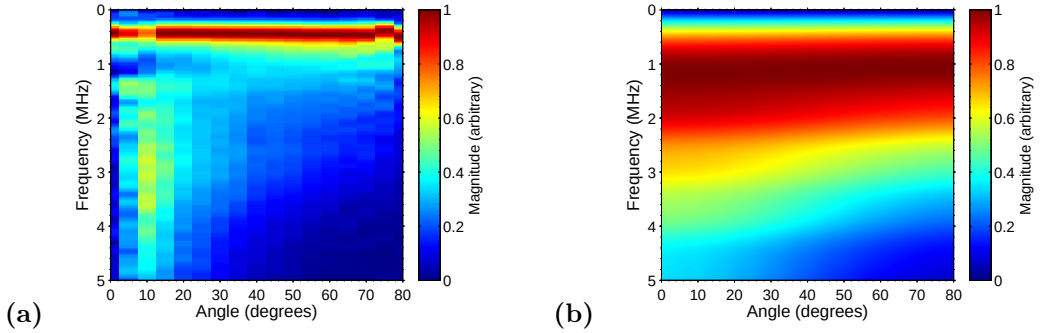


Figure 5.6: The experimentally measured (a) and simulated (b) frequency response for a 10 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

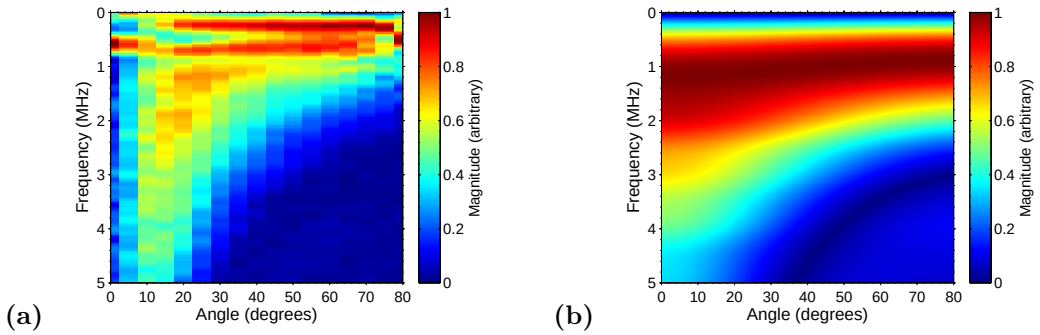


Figure 5.7: The experimentally measured (a) and simulated (b) frequency response for a 20 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

5.4 Frequency response

5.4.1 Variation with angle and coil size

The A-scan collected for each incident angle, using the set-up depicted in figure 5.3 with the laser line source acting ablatively, was converted to the frequency domain using an FFT, and the magnitude taken. The A-scans were of signals averaged from 16 collections, and these measurements were themselves repeated four times. The mean of the magnitudes for each incidence angle were taken, and then scaled such that for a given incident angle, the maximum magnitude was one and the minimum zero (no offset was added, zero was simply left as zero). Results were collected for the EMATs described in table 5.1. The EMATs were modelled using these same parameters and the method described previously, using a wave speed of 5932m/s; only compression waves were modelled, as the experimental results for shear waves were not of a sufficient standard to form a meaningful comparison with simulated results. The results comparing the experimental and simulated data are shown in figures 5.6 to 5.12.

The troughs in the frequency domain of the experimental and simulated data are the dark blue regions of figures 5.6 to 5.12. Unfortunately, the troughs in the experimental data are difficult to see, and various attempts to make them clearer by manipulating the colour scale were unsuccessful. However, under careful inspection, the agreement between experiment and simulation, regarding the position of the troughs, is good. This would suggest that the simulation is correct as far as

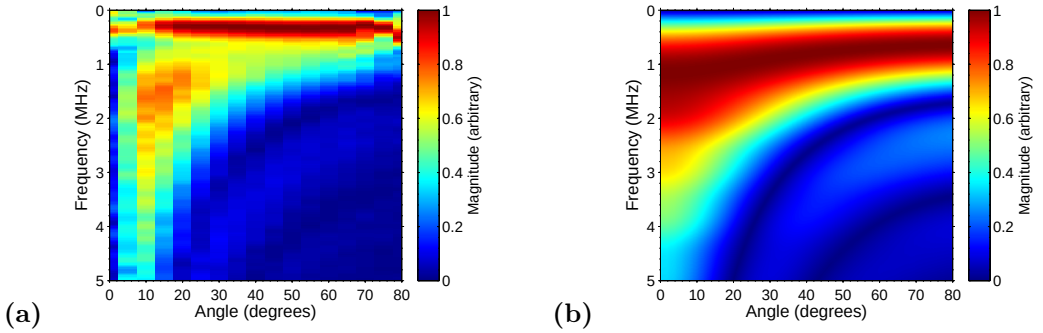


Figure 5.8: The experimentally measured (a) and simulated (b) frequency response for a 30 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

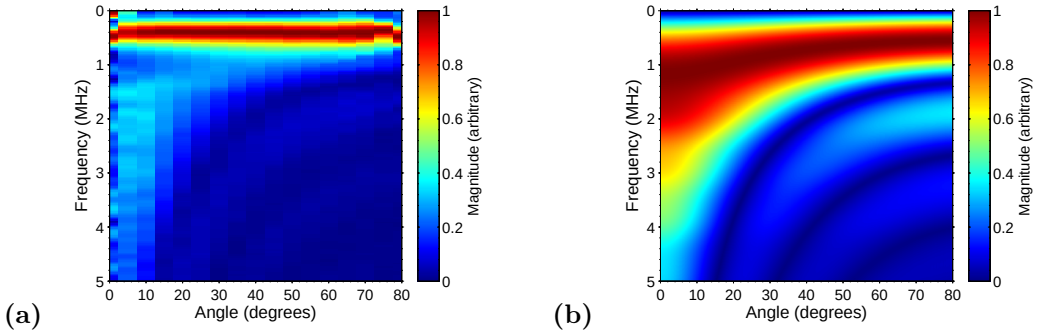


Figure 5.9: The experimentally measured (a) and simulated (b) frequency response for a 40 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

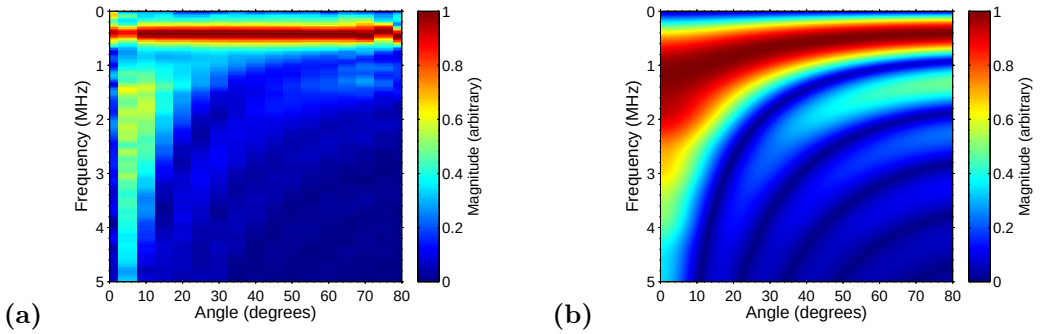


Figure 5.10: The experimentally measured (a) and simulated (b) frequency response for a 50 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

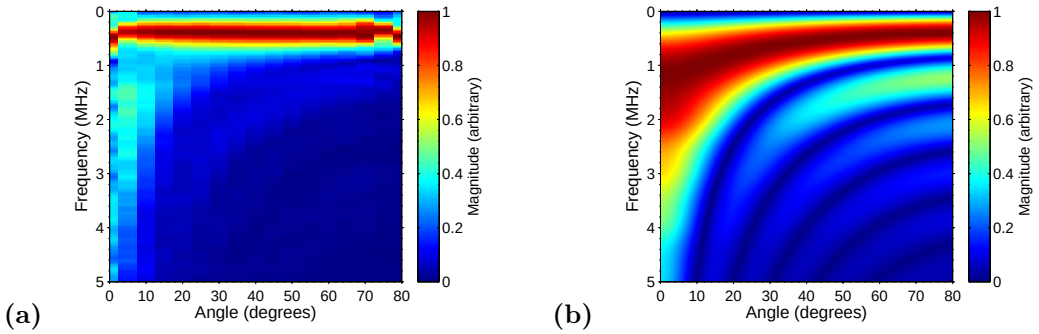


Figure 5.11: The experimentally measured (a) and simulated (b) frequency response for a 60 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

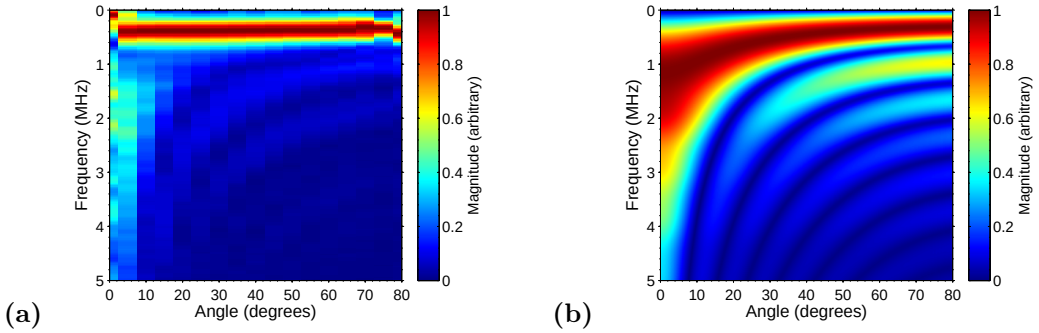


Figure 5.12: The experimentally measured (a) and simulated (b) frequency response for a 70 turn EMAT operating on a steel sample. The frequency magnitude has been scaled for each incident angle such that the maximum magnitude at a given angle is one, and the minimum zero.

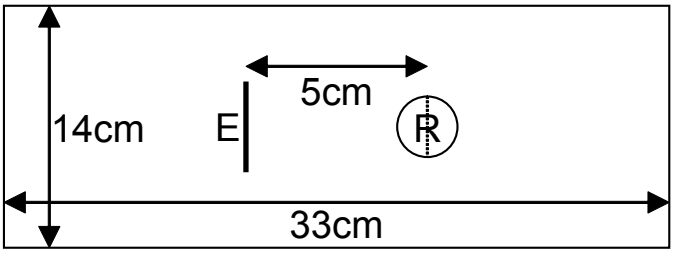


Figure 5.13: This is a top-down view of the experimental set-up used for EMAT SIR measurements. The EMAT is placed onto a flat and level block of mild steel, 33cm long (horizontally across the diagram), 14cm wide (vertically across the diagram), and 13.7cm deep (into the diagram). Relative to the centre of the top surface of the block, the EMAT was placed 2.5cm out horizontally from the centre, whilst staying in the centre of the width of the block; the coils were orientated along the width of the block (vertically across the diagram, as shown by the dashed line through the circle with an "R" inside). The laser line source, as used previously, was focused 2.5cm out horizontally from the centre, whilst staying in the centre of the width of the block; the line source was orientated along the width of the block (vertically across the diagram). Different coil widths were used to explore the frequency response due to variations in receiver size, whilst keeping other parameters constant wherever possible.

it describes the effect of the size of the transducer, but incomplete, as it does not include the frequency response due to the impedance of the coil, or other effects that vary with frequency. In addition, it would appear that taking the signal at a single incident angle for the 10 turn coil EMAT, and using it to simulate every case, was insufficient, as the overall shape of frequency response of the simulated data does not match the experimental data.

5.4.2 Variation with coil size for surface waves

The set-up depicted in figure 5.13 was used, with the laser line source acting ablatively; an A-scan was collected for each coil. Only the temporal region containing the Rayleigh wave was captured, using signals averaged from 16 collections. Results were collected for the EMATs described in table 5.1.

The response of the EMATs was modelled as previously described for a plane wave front; this is acceptable due to the use of a line source that is longer than the length of the EMAT coil being used as a receiver. Setting the angle to zero in equations 5.10 and 5.11:

$$F(\omega) = \frac{2Rv}{\omega} \sin\left(\frac{|\mathbf{s}|\,\omega}{2v}\right) \tag{5.15}$$

$$F(0) = R\,|\mathbf{s}| \tag{5.16}$$

The EMATs were modelled using the parameters in table 5.1 and a wave speed of 2987m/s. Only Rayleigh waves were modelled, as the experimental results for compression waves were not of a sufficient standard to form a meaningful comparison with simulated results. The frequency domain transfer function due to the size of the EMAT was calculated and can be seen in figure 5.14a. At each trough, a phase shift that is some multiple of 180° is expected. To test this, the experimental A-scans were converted to the frequency domain using an FFT, and the phase extracted. The phase was then unwrapped, and the linear trend removed, such that large jumps in phase could easily be seen (figure 5.14b). The phase jumps are in approximately the correct place

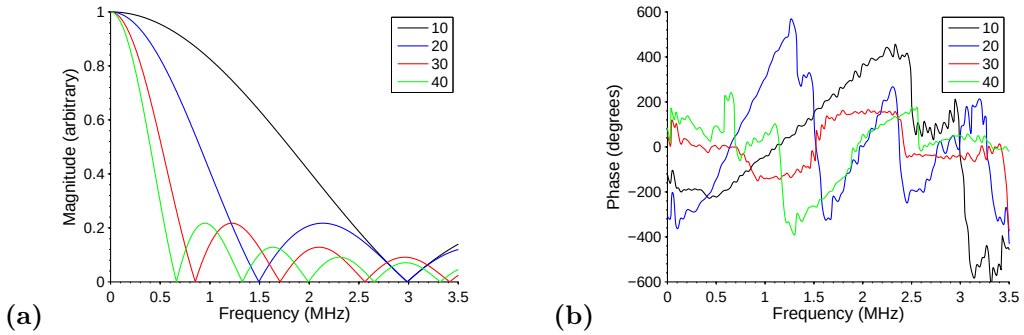


Figure 5.14: The simulated frequency spectrum magnitude (a) and experimentally measured phase at each frequency (b) for a series of linear EMAT coils, with 10, 20, 30, and 40 turns (as noted in the legend), operating on a mild steel sample. The frequency spectrum magnitude has been scaled such that the maximum magnitude is one. Where the simulated frequency spectrum magnitude nulls are found (a), a large change in phase is observed in the experimental data (b), which is consistent with a null being present in the experimental frequency spectrum magnitude. Considering the phase is the clearest way of observing these nulls in the experimental data.

in terms of frequency, and approximately a multiple of 180°, although unfortunately the results cannot be claimed to be particularly clear.

It is necessary to check the transforms calculated due to the EMAT coil spatial width are accurate. The A-scans were converted to the frequency domain using an FFT, and the magnitude taken. These magnitudes are scaled such that for a given coil, the maximum magnitude was one and the minimum zero (no offset was added, zero was simply left as zero), forming figure 5.15a. To compare with the calculated transforms, the experimentally measured frequency domain form for the coil with 10 turns was divided by the calculated response for a 10 turn EMAT (such as the response shown in figure 5.14a). The result of this operation was then multiplied by the calculated transforms (figure 5.14a) for EMAT coils of the other number of turns that are under test, to produce an expected frequency domain response for each coil, 5.15b. The output of the transform was set to zero for any frequencies above the first trough for each calculated response. Directly comparing figures 5.15a and 5.15b, it can be seen that the frequency of the first trough has good agreement between the experimental and simulated data. This would strongly suggest that the model used was valid for this purpose.

The model can be used to calculate the upper useful frequency for any linear EMAT coil used. However, it cannot predict the form of the signal received, nor the peak-to-peak amplitude, and hence its application is limited.

5.5 Conclusions

Any transducer has a finite size for both generation and detection, and the affect of non-zero size on frequency response must be considered. Models were constructed for both far-field and near-field cases. Simulated data was found to agree with experimental tests when considering nulls in the frequency response. The models were too narrow in scope to fully describe the impact of changing coil configuration on the frequency response; for example, they do not include frequency changes due to changes in mutual and self inductance. However, it is beneficial to explore the change just due to transducer size, and that is what this work has achieved. It could potentially be combined

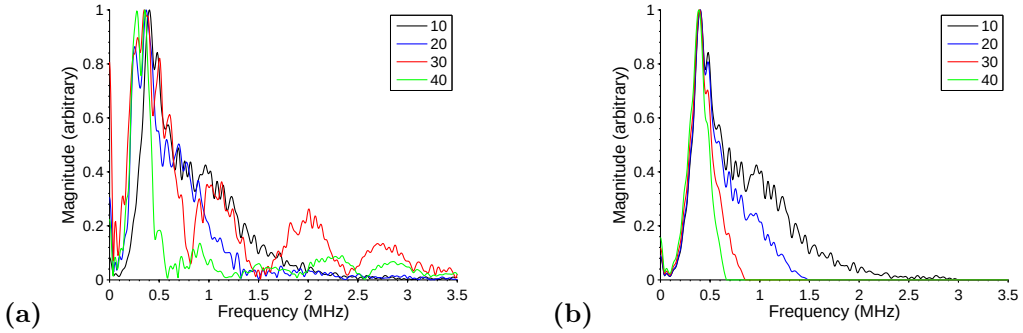


Figure 5.15: The experimentally measured (a) and simulated (b) frequency spectrum for a series of linear EMAT coils, with 10, 20, 30, and 40 turns (as noted in the legend), operating on a mild steel sample. The simulated results were obtained by dividing the frequency domain form for the EMAT coil with 10 turns by the calculated response for a 10 turn EMAT coil, and then multiplying by the calculated transforms for EMAT coils of the other number of turns that are under test. The output of the transform was scaled such that the maximum magnitude is one, and set to zero for any frequencies above the first trough for each calculated response.

with models of the other factors impacting on EMAT response in future work.

5.6 Use in development of TOFDI

This chapter has compared a simple model of how the spatial impulse response of a linear EMAT coil is related to coil width, against experimental data. Although insufficient factors have been included to provide a matching waveform, the model is sufficiently quantitative to correctly predict the troughs in the frequency domain, and the variation in the trough position as the ultrasound angle of incidence changes. This information can be used to choose a linear coil that is most suitable for use in TOFDI, as implemented within this work. From the model and the experimental data, it can be seen that the EMATs retain sufficient frequency bandwidth (80% of peak amplitude up to 1MHz at 80°, this frequency range containing most of the signal energy) at coil widths equivalent to 30 turns. This makes 10, 20, and 30 turn coils viable. Although not presented here, data on peak-to-peak amplitude variation with angle, and data on peak-to-peak amplitude variation with lift-off, have been collected. The 30 turn coil did not have sufficient peak-to-peak amplitude relative to the 10 and 20 turn coils, and thus has been disregarded. The 10 turn coil can offer superior performance if there is zero lift-off, but above this, it drops to the same level as a 20 turn coil. Since the drop is so much less for a 20 turn coil (although it is still a significant drop in peak-to-peak amplitude with lift-off), it was chosen instead of the 10 turn coil, as it is expected that the lift-off will vary when performing a TOFDI scan, if TOFDI is being used for the prototypical application. Less variation in peak-to-peak amplitude due to lift-off, leads to a more consistent signal, which then requires less signal processing stages.

Chapter 6

Removing direct and reflected waves

This chapter has been published [165] in a modified form in a peer-reviewed journal.

This chapter describes a method for removing ultrasonic waves from B-scans, that have either travelled directly from the emitter over the surface to the receiver, or that have been reflected from the back-wall, or that have gone from the emitter, along the surface to the sample edge, and back to the receiver. These waves do not contain useful information on bulk defects, but they can obscure bulk waves scattered by defects, and they can also make examining the B-scan more difficult, as their presence means that much of the dynamic range of the B-scan is occupied with these very strong directly received or reflected waves, rather than the weaker scattered waves. As the cross-sectional imaging stage of TOFDI requires the defect indications to not be heavily obscured by coherent interference, and B-scans are also shown as part of TOFDI, a technique to remove these lines is required.

6.1 Introduction

The Hough transform (HT) [166] is an effective method to detect straight lines or other parameterisable shapes [167], and is known to be robust to noise [168]. The HT is used in areas including machine vision [169], electron back-scatter diffraction (EBSD) analysis [170–172], and feature location in geoscientific images [173]. In NDT, it can locate parabolas in B-scans [17,18,34], although such work often uses an inverse Hough transform (IHT) [174] and genetic algorithms to solve the consequent parameter optimisation problem [18,34].

Due to image discretisation, lines in the image space can be split into "thick" (figure 6.1a) and "thin" (figure 6.1b) lines [175,176], where a thick line is split into a series of vertical or horizontal segments. It is clear from this that a thin line can only exist at 45° (in either direction), and that the extremes of thick lines are found at 0° and 90°. However, the issue of "sparse" lines does not appear to have been considered previously.

6.1.1 TOFD and B-scan interference

TOFD B-scans, with scan position and time forming the horizontal and vertical B-scan axes respectively, show the standard signals, which are the direct, back-wall reflected, and defect scattered waves incident on the receiver, but in addition, they can show a wave that has gone from the emitter, along the surface to the sample edge, and back to the receiver (figure 6.2). Such waves form

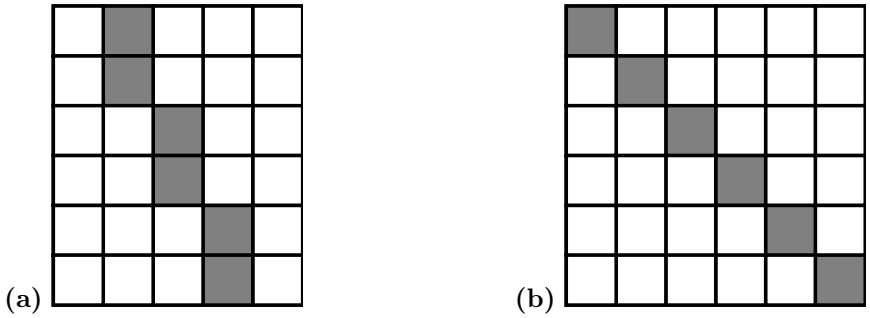


Figure 6.1: Image (a) shows a "thick" line; a single row/column can contain multiple line elements. Image (b) shows a "thin" line, where each column/row of the image array can only contain a single element of the line.

angled lines in the B-scan, and can cause significant interference to any parabolas in the B-scan which are created by bulk waves scattering off defects (both compression and shear waves can be used in TOFD [48]). Even where the surface and bulk waves are not overlapping, the presence of such a strong signal in the B-scan can introduce artefacts into the automated analysis if time-gating cannot completely remove them (if the surface waves are between diffracted signal returns from two defects for example). Removal of interfering waves is complicated by the issue that the interference is not necessarily constructive, hence two waves could destructively interfere to create a null at a certain point in the B-scan, but not interfere at another point in the B-scan.

6.1.2 Thick lines in B-scans

Consider an idealised signal, a Dirac delta function, time-shifted by the receiver-emitter separation, r, divided by the wave speed, c:

$$s(t) = \delta(t - r/c) \tag{6.1}$$

If emitter and receiver are in a fixed relative geometry during a B-scan, as shown in figure 6.2, waves travelling directly from emitter to receiver have a constant time shift during the scan, forming horizontal lines in the B-scan. Considering the common case of the scan motion, Δr, being parallel to the line between emitter and receiver, if a reflecting edge is perpendicular to the emitter-receiver line, waves will reflect off the edge and return to the receiver in the same direction (r_0 is the initial distance from emitter to edge and back to receiver):

$$s(t) = \delta(t - (r_0 + 2\Delta r)/c) \tag{6.2}$$

If the time difference is such that the times of arrival do not fill adjacent (including diagonally) discretely sampled time points, the line will be "sparse" rather than "thick" as the time segments between the two points will not be filled (figure 6.3a). This could be considered a form of aliasing, although usually when discussing aliasing of images, both axes of the image represent spatial coordinates, whereas here one axes is spatial, and the other is temporal. In this case, the spatial sampling period, Δr, required to avoid this issue is (T_s is the temporal sampling period assuming the idealised signal fills just one sample):

$$\Delta r \leq cT_s/2 \tag{6.3}$$

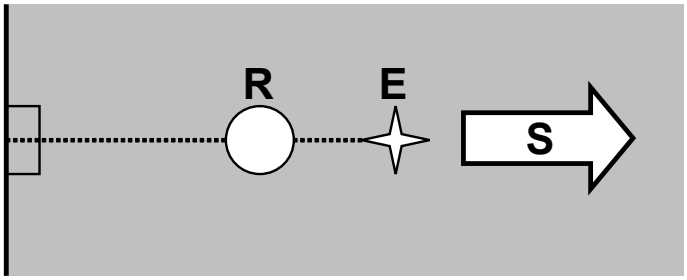


Figure 6.2: The most common geometry (not the only possible one) for a reflecting edge in a TOFD scenario has emitter, E, and receiver, R, along a common line with the scanning direction, S, and a reflecting edge perpendicular to this line. This geometry will produce a linear reflected line in the B-scan as simulated in figure 1.1b.

A Rayleigh wave has a typical speed of 2996m/s for mild steel [19]. Sampling an ideal signal at a sampling frequency of 10MHz leads to a Δr of 0.15mm. Previous NDT work has tended to use steps in the 1mm range [10]. It is preferable to have more freedom over T_s and Δr, accepting that modifications to the HT will be required. This is unlikely to be of significance for parabolas due to the slower change in time of arrival with scan position.

For low bandwidth, temporally wide signals, the amplitude variation between scan positions may be small (figure 6.3b), but this is not the case for temporally sharp signals used in techniques such as TOFDI [163]. In actual usage, the problem is largely invariant with sampling frequency, as sampling frequencies of 100MHz or 1GHz are easily achievable. Consequently, the signal will never appear as ideal, instead appearing as a band, but higher sampling rates will make any gaps larger in terms of the number of sample points. The primary issue of concern is then how rapidly the reflected signal varies. Consider a Δr of 0.8mm and a wave speed of 2996m/s. If the reflected pulse varies significantly over 0.5μs, the line would appear sparse. Piezoelectric transducers can readily reach pulse durations of 0.5μs, but since the objective of this work is to remove such lines from images leaving underlying signals intact, the pulse duration is of less interest than the pulse frequency. To subtract the line correctly requires operation on the signal prior to passing through envelope detection, such that all operations on the underlying signal would proceed correctly after the subtraction has been performed. Consequently, a pulse with a centre frequency of only 1MHz would result in an amplitude that varied from its maximum value to zero in just 0.25μs, neglecting the additional effects of the pulse window. The problem would worsen as the pulse frequency increased or if Δr increased, making lines sharper and steeper respectively.

6.2 Modified Hough transform

Hough [166] originally used $y = mx+b$ to transform from image space to Hough space, but this had the problem that m was unbounded and $m = \infty$ for vertical lines. Instead, the discretised image space (x, y) can be transformed to a discrete Hough space using the "normal" parameterisation (θ, ρ), as shown in figure 6.4, converting lines to points [167] in a one-to-many mapping. This changes the problem from the difficult search for lines, to the simpler search for peaks generated by accumulation (known as voting in the literature), where the sinusoidal lines that each point in image space makes in Hough space overlap (corresponding to collinear figure points) [167].

The parametrised line is normal to the line from the origin at angle θ with respect to the x-axis,

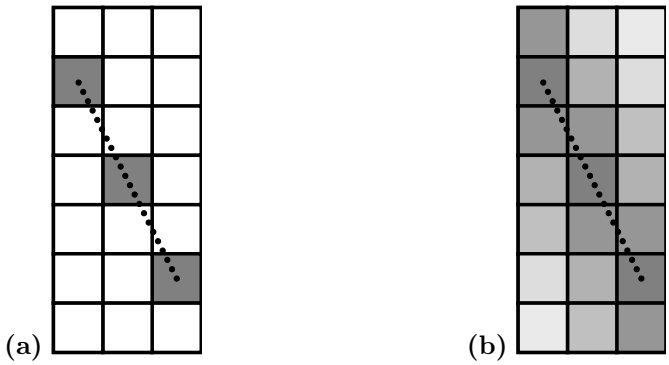


Figure 6.3: Image (a) depicts a "sparse" line, split into short vertical segments. Each image column can only contain a single element of the line and would be observed in a B-scan for any single pixel width line with an angle more vertical than 45°. Image (b) depicts the same line with longer vertical segments; sparsity is concealed by the slow variation of pixel amplitude.

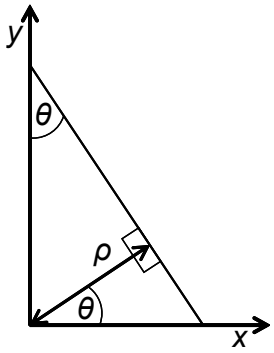


Figure 6.4: Normal parameterisation of a straight line, used to transform from space to Hough space.

and intersects at a distance ρ from the origin, which is set at image centre for convenience [173]. The acceptable error in θ and ρ is used to quantise Hough space:

$$\rho = x\cos\theta + y\sin\theta \tag{6.4}$$

The normal parameterisation can be reached from the standard $y = mx + b$ by the simple substitution:

$$x = \rho\cos\theta \tag{6.5}$$

$$y = \rho\sin\theta \tag{6.6}$$

$$\rho^2 = x^2 + y^2 \tag{6.7}$$

The HT is not designed for sparse lines, but scaling the HT by the steepest line gradient (Δy_{max}) which will be encountered, transforms the steepest angle to 45° (from <45°). Consequently, the lines will not appear sparse to the HT, as for each vertical pixel change, there is at least a single horizontal pixel change. This assumes there are no purely vertical lines present such that every angle is >0°. From the normal parameterisation:

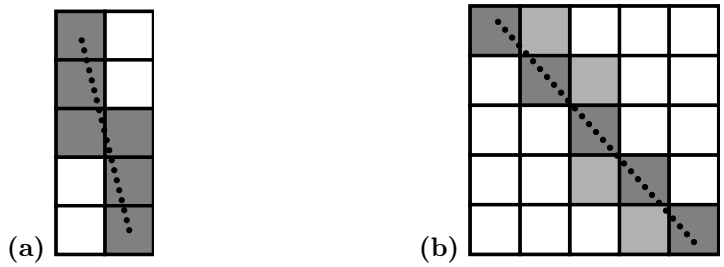


Figure 6.5: Non-sparse lines with thickness along both axes. Image (a) has $\theta = 21.8°$ and is primarily vertically thick and secondarily horizontally thick. Image (b) has $\theta = 45°$ but ρ is such that the line does not start at a pixel centre, and thickness occurs without a primary direction (pixels selected for a ρ starting at pixel centre are shown darker than the other selected pixels).

$$\tan\theta = \frac{\Delta x}{\Delta y} \tag{6.8}$$

Consider a change in the x-axis, $\Delta x_{min} = 1$, in pixel units. If the largest change in y (smallest angle of interest) is Δy_{max} (not necessarily y_{max}):

$$\tan\theta_{min} = 1/\Delta y_{max} \tag{6.9}$$

It is conceptually simpler to make the x-axis larger than the y-axis smaller since this allows the use of integer pixel positions (unnecessary from an implementation perspective). Therefore the x-axis is scaled so that $\theta_{min} \to 45°$ and $45° \leq \theta \leq 135°$:

$$x' = x\Delta y_{max} \tag{6.10}$$

$$\tan\theta' = \frac{\Delta x}{\Delta y}\Delta y_{max} \tag{6.11}$$

The HT may still incorporate very mild thickness along either axis due to the issue of the line crossing at a pixel edge rather than a pixel corner (figure 6.5). This is too minor an issue to be of concern, involving only single additional pixels per column or row, as opposed to the numerous pixels of thickness that can occur with actual thick lines. It is alleviated by most lines not varying significantly over just two pixels, since the width is often just very narrow rather than a single pixel. For lines of any extent, this issue does not occur at 45° or 135° for ρ values which place the line such that it passes exactly through the centre of each pixel.

Once points in the Hough space have been identified using a threshold or otherwise, the IHT can be used to show where the selected lines appear in the image space:

$$y = \frac{\rho - x\cos\theta}{\sin\theta} \tag{6.12}$$

In this form, vertical lines only exist at a single x value, and thick lines in the y-direction require x to be oversampled, so that if $y_1, \ldots, y_n$ are all at the same discrete x_m point, then there will need to be at least n points in the x range $\frac{x_{m-1}+x_m}{2} \leq x_m < \frac{x_m+x_{m+1}}{2}$ in order to avoid gaps in the line. Oversampling is not usually a problem as the IHT uses far less memory and computation than the HT, as only a small selection of the pixels in Hough space need to be transformed (those identified at the earlier stage). Alternatively, the IHT can be performed in two passes, once with y

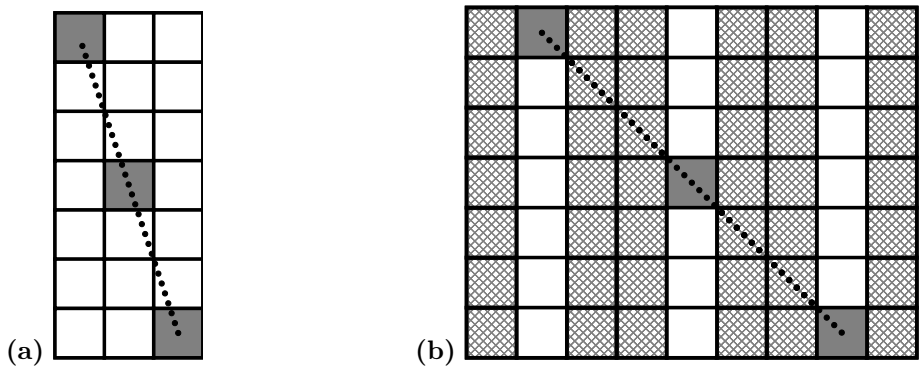


Figure 6.6: Image (a) depicts a sparse line, (b) depicts the same line after x-axis transformation. Scaling by $\Delta y_{max} = 3$, the line angle transforms from 18° to 45°. The dotted line only crosses one "real" pixel per column and the HT will operate as if the line were continuous. Cross-hatched pixels are not real in the sense that they are a mathematical convenience and contain no data.

the subject, for thin lines and lines that are thick in the x-direction, and once with x the subject, for thin lines and lines that are thick in the y-direction. However, now the scaled values must also be used in the IHT (ρ' is ρ for the scaled axes):

$$y' = \frac{\rho' - x' \cos\theta'}{\sin\theta'} \tag{6.13}$$

In this form, there are no vertical lines, and oversampling of the IHT should not be required as there are no thick lines in the y-direction. If oversampling is used, it should be noted that the IHT does not progress smoothly between the discrete values of x', as it was assumed in the HT that the values in-between contain no data (figure 6.6). Instead, the oversampling is, for each discretised x' value (x'_{mid}), in the range $(x'_{mid} - 0.5) \leq x' < (x'_{mid} + 0.5)$. This assumes an initial image pixel width of unity with a centre at x'_{mid}, and oversamples over that pixel width, but not outside the pixel. Some techniques use the IHT to perform the voting in the image space rather than the Hough space [18,34], but this is not done here; the IHT is used only to draw lines based on voting by the HT. It should also be noted that the angle and offset of a line detected after transformation is equal to the following pre-transformation values:

$$\theta = \arctan\left(\frac{x_{max}}{x'_{max}} \tan\theta'\right) \tag{6.14}$$

$$\rho = \sqrt{\left(\frac{x_{max}}{x'_{max}} \rho' \cos\theta'\right)^2 + (\rho' \sin\theta')^2} \tag{6.15}$$

Scaling the axes means that two vertically adjacent pixels should have no overlapping points in Hough space. This is true in a continuous Hough space, but in a discrete Hough space there could be overlap due to the discretisation process. The smallest change in ρ with y will occur at the extremes of the picture:

$$\Delta\rho_{min} = \sqrt{x_{max}^2 + y_{max}^2} - \sqrt{x_{max}^2 + (y_{max} - 1)^2} \tag{6.16}$$

Setting $\Delta\rho$ to $\Delta\rho_{min}$ (the ρ increment required for there to be no overlap for vertically adjacent pixels anywhere in the image) often results in a Hough space that is far too finely discretised to fit memory and speed of computation requirements. Hence, $\Delta\rho_{min}$ is taken as a lower bound and

$\Delta\rho = 1$ as a default. The slight overlap in Hough space of vertically adjacent pixels transformed from the image extremes is only of minor concern. In addition, it is sensible to ensure that $\frac{1}{\Delta\rho} \in \mathbb{Z}^+$ so that integer values of ρ are reached after $\frac{1}{\Delta\rho}$ steps.

The most vertical angle that is still useful corresponds to a line which extends from 1 to y_{max}, and only changes by a single pixel in the x dimension at the boundary from y_{max} to $y_{max} + 1$:

$$\tan\theta_{min} = 1/2y_{max} \tag{6.17}$$

Using θ_{min} usually results in a Hough space that is scaled to the extent that it is too large to fit memory and speed of computation requirements, and usually a much larger angle can be used based on what is the steepest expected line in practical usage. Similarly, there is little point in having an angle increment $\Delta\theta < \theta_{min}$ as there will be significant duplication in the Hough space (multiple points will represent the same line), and practical experience demonstrates that a larger angle increment still provides good performance.

Note that although the maximum useful range for ρ has been defined for a central origin, this assumes a continuous distribution of angles between and including the limits. Equivalent to the $|\rho_{max}| = \sqrt{x_{max}^2 + y_{max}^2}$ formulation, the maximum ρ occurs at $\frac{\partial\rho}{\partial\theta} = 0$, and is given by:

$$\theta_{\rho_{max}} = \arctan\left(\frac{y_{max}}{x_{max}}\right) \tag{6.18}$$

$$|\rho_{max}| = x_{max}\cos\theta_{\rho_{max}} + y_{max}\sin\theta_{\rho_{max}} = -x_{max}\cos\left(\pi - \theta_{\rho_{max}}\right) + y_{max}\sin\left(\pi - \theta_{\rho_{max}}\right) \tag{6.19}$$

Where $-y_{max} \leq y \leq y_{max}$ and $-x_{max} \leq x \leq x_{max}$. If this angle is never hit exactly, then the ρ_{max} is never reached, and the effective maximum can be approximated as the ρ value at x_{max} and y_{max}, when θ is nearest $\theta_{\rho_{max}}$. This is only useful if a very restricted range of angles is being used and memory usage must be minimised by minimising the ρ range.

The HT was originally designed to deal with binary lines of pixel width and each vote contributed a fixed value [167]. Later, a vote was made equivalent to the intensity value at that point in the image [169]. In contrast, B-scans contain bands of varying amplitude. Bands of varying magnitude are dealt with in EBSD analysis [170–172], with early automated analysis finding edges by grouping pixels into regions based on gradient orientation correlations [170], but these bands appear to have a slowly varying magnitude between their edges, and then a rapid variation in magnitude at the edge, whereas bands in a B-scan do not have clear edges, and can drop off relatively slowly at their temporal extremes. In addition, the increased intensity between the edges of an EBSD band can be used to identify the centre [172], whereas bands in B-scans may be asymmetric, without a clear centre. Consequently, edge detection methods, such as Roberts Cross, Sobel, Hueckel and the Canny edge detector [177], have been avoided at this stage, and instead the bands are treated as a series of parallel lines, taking care to avoid spurious identification of lines with angles close to the correct angle. In order to deal with this, a modified detection criterion and θ grouping are required, and to facilitate the former, more information is required from the transform. First, votes are scaled by the originating values of the points in the image and the accumulation, $\sum A$, is output. The scaling value can be positive or negative, if amplitude, A, rather than magnitude, $|A|$, has been input.

Next, the number of votes added to each point in Hough space, N, (for normalisation, $\bar{A} =$

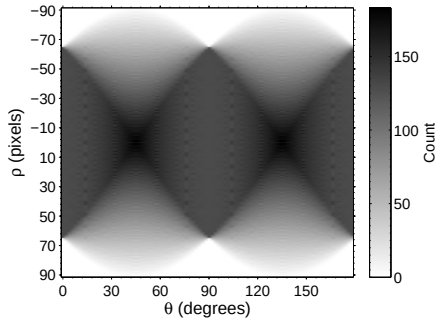


Figure 6.7: The spread of counts in Hough space for a uniform input space. The input is 128 pixels square, the angles are $0° \leq \theta < 180°$ with $\Delta\theta = 2°$, the perpendicular distances are $-90 \leq \rho \leq 90$ with $\Delta\rho = 1$.

$\frac{1}{N}\sum A$) is output, since for a completely uniform image, the HT is not uniform [17,172]; longer lines accumulate more points and will appear more prominently despite an equivalent mean magnitude. The literature refers to this as biasing or as a windowing effect, and it can be eliminated by dividing by the transform of the aforementioned completely uniform image, at the risk of enhancing the contribution of noise in the less populated parts of the Hough space. Other workers have used a factor to "soften" the normalisation [173], to ensure that very short lines are unlikely to go above the detection threshold just because of noise, and to still give long lines some priority. Alternatively, the image can be made into a circular array [171] so that the biasing does not contain large intensity gradients, but this is not a possibility here as it would remove far too much useful information from the B-scan. This work is less concerned with very short lines, as false detection and removal of such lines has only a very small affect on the overall B-scan. This allows lines of only a few pixels to be ignored completely due to their insignificance to the overall B-scan, and consequently, a simple normalisation from dividing by the spread of counts (figure 6.7) is viable.

The maximum count from figure 6.7 is not 128 pixels as would be expected. The simplest reason for this is coarse sampling of the Hough space causing multiple lines of similar (θ, ρ) to appear at the same discretised point in Hough space.

Also output is the standard deviation for each point in Hough space, σ (using $N-1$ rather than N, as the estimate of the mean results in the data lying closer to that mean than the population mean, and reduces the degrees of freedom by one):

$$\sigma = \sqrt{\frac{1}{N-1}\left[\left(\sum_{i=1}^{N} A_i^2\right) - N\bar{A}^2\right]} \tag{6.20}$$

Thresholding selects peaks in the Hough space, global thresholding being common [178], which is the choice here. The HT usually thresholds against the mean, $\bar{A}$ [17, 172]. Instead, consider the thresholding parameter γ, selecting lines on the criterion:

$$\sigma < \bar{A}\gamma \tag{6.21}$$

Therefore γ detects lines based on if their standard deviation is below a certain fraction of the mean, and such a criterion can detect greyscale lines successfully in the presence of an interfering greyscale background. Using the standard deviation has the advantage that if two lines have no overlapping parts, even if the lines have very different magnitudes (figure 6.8a) they should both

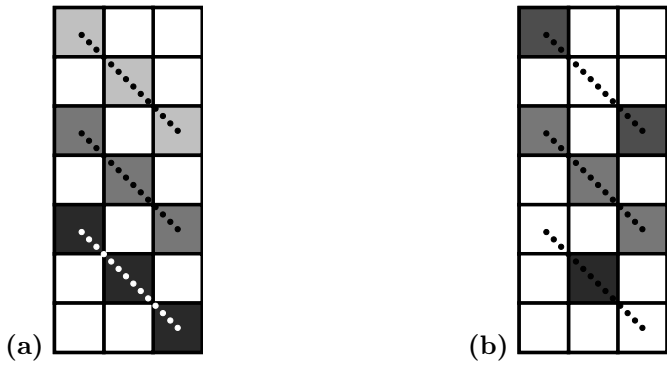


Figure 6.8: Depicting lines at 45° with no interference from other lines or noise, image (a) lines have the same standard deviation but different means, and image (b) lines have the same mean but different standard deviations.

appear equal in the standard deviation plot but will not in the mean plot. Noise will have a greater affect on the standard deviation of the lower mean signal, assuming that the noise distribution is not dependent on the local amplitude, and this is compensated for by scaling the threshold by the mean. Hence strong signals in noisy situations will be selected in preference to weak signals in noisy situations (where miss-identification of random noise as a line is more likely), but in very low noise conditions both sets of lines will be selected. Points in Hough space that have had less than three points added ($N < 3$) are removed from consideration. In addition, the mean is unable to distinguish a complete line of amplitude A from a line with only a third of the points but amplitude $3A$ (figure 6.8b), whereas the standard deviation will show a marked difference. This is an important property when looking for lines of low amplitude near non-intersecting lines of high amplitude, as setting a threshold is complicated when the mean of three filled identical points is equal to the mean of three points, only one of which has a non-zero (and three times larger) amplitude. Using the mean would find a spurious line of mostly zeroes, that just happened to cross a strong line at some point, as readily as a real line of low but constant amplitude.

The standard deviation is also useful as it makes it easier to find a large number of adjacent parallel lines of different amplitude without accidental identification of the "band" as a series of lines at slightly different angles. This is because the mean is expected to vary within a band, but the standard deviation should be approximately constant (assuming little interference).

The point spread function (PSF, the appearance that a single line in image space has in Hough space) has a characteristic butterfly shape for the normal parameterisation [171, 175, 176]. A peak in Hough space spreads in the ρ direction as $\Delta\rho$ decreases and spreads in the θ direction as $\Delta\theta$ decreases [176]. "Butterfly" masks have been developed to enhance the peaks [171, 172] but these are not used here. After thresholding has selected the peaks in the Hough space, grouping of the θ parameter (ρ and xy grouping [179] is not used here) identifies parallel lines. For this work, grouping consists of generating a histogram of lines at each θ and searching for local minima around the maxima in order to split the groups. Large groups suggest the presence of a band at the maximum histogram peak in that group. That angle can then be picked, excluding other angles in that histogram group. This reduces identification of spurious lines at angles close to the true angle, and allows for a lower global threshold in order to identify lines that are just off the angle tested, but does assume that there are not important lines actually at those adjacent angles rather than at the selected angle. Once grouping selects the "true" peaks, another HT can be performed, if necessary, only using (θ, ρ) parameters around those peaks, but at much higher

parameter resolution. Known as iterative focusing [178], the Hough space is initially coarse but uniform, and then it is iteratively focused to higher resolutions around peaks of interest, but this technique is not used here. When using a HT for EBSD, poor ρ localisation may occur due to the lines actually being bands [172]. This is not a problem in the ultrasound B-scan case since it is desirable to fit a line for each pixel width of a band, assuming the correct angle has been found, as the lines are assumed to be of pixel width. Poor θ localisation can lead to problems removing the lines, hence the importance of the grouping stage, especially for shorter lines.

The standard Hough transform does not concern itself with contiguity [167], although there are plenty of variations of the Hough transform that attempt to localise the line segments [169, 175, 176, 179] if contiguity cannot be assumed. Lines caused by direct and reflected surface waves will extend from one edge of the B-scan to the other under most circumstances, entering the B-scan at a certain scan position and time (with one of these parameters being the maximum or minimum for the range shown by the B-scan), and leaving the B-scan at another scan position and time (with the other parameter being the maximum or minimum for the range shown by the B-scan). This is in contrast to a scattered bulk wave, which is more likely to only exist in a localised region of the B-scan that will not usually extend between the edges of a B-scan (unless the scan is over a very restricted positional range, or the tails of the scattered signal in the B-scan are very long). Due to this, localisation of lines to finite lengths would offer no benefit at this early stage (this is also the case for EBSD [170, 172]).

This modified HT has been applied to the detection and removal of angled lines in a TOFDI B-scan [163].

6.3 Usage of the modified Hough transform

First the data is band-pass filtered to reduce noise. Then, the B-scan amplitude (not magnitude) data is passed into the HT, as this allows interfering waves to be correctly handled. Afterwards, the prior band-pass filter is applied again as removal of lines can introduce frequencies that are outside of the original signal frequencies.

6.3.1 Direct surface waves

Direct surface waves are always present and have a high amplitude, and so are tackled separately and first. Direct surface waves can especially cause problems by obscuring echo data that arrives very soon after emission [10]. Since these waves are at a single angle in the B-scan (90°) coincident with the x-axis (the scan position), the median of each horizontal line is subtracted from the line. Then, on an individual pixel basis along each line, any magnitudes greater than the original pixel magnitude are forced to the original pixel magnitude, but the sign remains as that of the subtracted value.

Consider a series of horizontal lines of increasing amplitude contaminated by AWGN, as depicted by figure 6.9a. The increasing amplitude of the horizontal lines manifests as a darkening gradient down the image as each horizontal line gets a progressively higher amplitude. The ideal way to remove the lines in this case would be to subtract the mean or median along each horizontal line, leaving only the noise. These horizontal lines represent surface waves, and where direct surface waves are strong, this subtraction removes them very effectively. This is because most interfering signals will be much weaker and will not exist along the entire line, and hence are averaged out.

Now consider the case of a weak horizontal line with a strong interfering angled line crossing it. Figure 6.10a represents this by treating a series of horizontal lines of the same amplitude as a uniform background; each horizontal line of the background is the same. The vertical line in figure 6.10a represents an angled line in a B-scan, which is crossing each of the horizontal lines that makes up the background of the scan. The amplitude of the vertical line varies, such that each horizontal line can be considered to be crossed by an angled line of a different amplitude. Despite being narrow, at the bottom of the figure the relatively high amplitude of the vertical line results in it having a much higher mean along the horizontal direction than the horizontal line, and subtracting the mean of each horizontal line rather than the median results in a DC offset. For example, consider if the background had an amplitude of 1 at each pixel, and each horizontal line had 10 pixels. If the vertical line was a single pixel wide, and had an amplitude of 2, then the median along a horizontal segment of the image would be 1 and the mean would be 1.1 (calculated from $(1 \times 2 + 9 \times 1)/10$), an increase of only 10%, with the median being the correct value to subtract in order to remove the background without a DC offset. If the vertical line had an amplitude of 10, the median would remain at 1, whereas the mean would be 1.9 (calculated from $(1 \times 10 + 9 \times 1)/10$). Now the increase is 90%, and again of course the median is the correct value to subtract in order to remove the background without a DC offset.

The magnitude (magnitude in this case is the amplitude without the sign) of any point along a horizontal line, due to the subtraction of the median, is not allowed to be larger than the original value, minimising the impact of strongly interfering signals increasing the median or signal gaps from being made into strong signals due to the subtraction. Such a restriction on the subtraction of the median minimises banding whilst still allowing other signals superimposed on the direct surface wave to appear after removal. Figure 6.10b again depicts horizontal lines (as in figure 6.10a), but now with a gap or a signal destructively superimposed, which the vertical line represents and hence has a reduced amplitude relative to the background. Removal of the background is still desirable, but not at the expense of creating a larger problem; subtracting the median without limiting the magnitude range of the result will lead to a large amplitude at a gap. With such a limitation in place, after subtraction the values along the vertical line would be the negative of the line until the line value reached below zero. After this point, the value at the line would simply equal the current line value. This causes a relatively small error if the line is a destructively interfering signal, but prevents a much larger error if the line is an attenuation of the horizontal line at that point. For example, consider if the background had an amplitude of 10 at each pixel, and each horizontal line had 10 pixels. If the vertical line was a single pixel wide, and had an amplitude of 1 (due to attenuation of the horizontal line at that point), then the median along a horizontal segment of the image would be 10. If this were subtracted, the vertical line pixel would now have an amplitude of -9, effectively creating a signal where none should exist. With the restriction in place, the pixel is limited to an amplitude in the range 1 to -1, and takes the value -1 in this case, a minor error from the correct result of 0. If the result should actually be -9 due to the vertical line representing a destructively interfering signal instead of an attenuation of the background, the error is clearly larger, and the vertical line pixel will effectively have been attenuated by the processing. There is no error if a vertical line is constructively interfering as in the previous case (figure 6.10a). It is considered better to lose signals in the small number of cases where a destructively interfering signal (vertical line in this case) has a consistently smaller amplitude than a background signal (horizontal line in this case) over the length of the vertical line, than to create signals where the background is attenuated. Without performing the processing described, it would not be possible

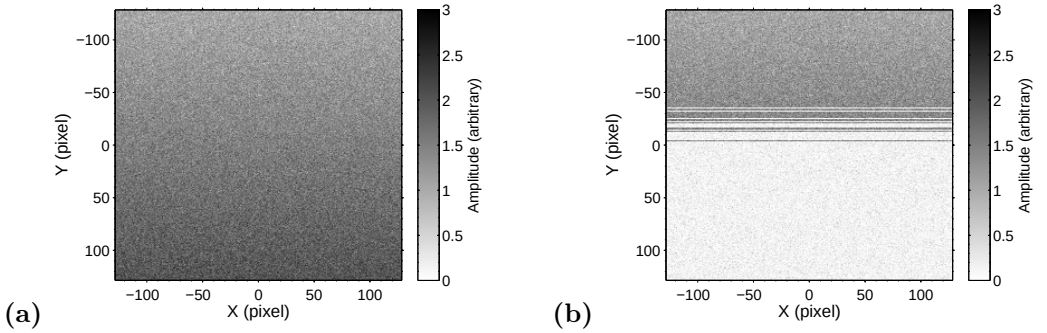


Figure 6.9: Image (a) has a vertical intensity gradient, A, linearly increasing from $A = 1$ to $A = 2$. AWGN has been added with $\overline{x} = 0$ and $\sigma = 0.2$. Image (b) has the mean subtracted for lines above the threshold ($\gamma = 0.14$). Thresholding based on the median would result in a similar problem in this case. The mean and the median are identical except due to the influence of noise.

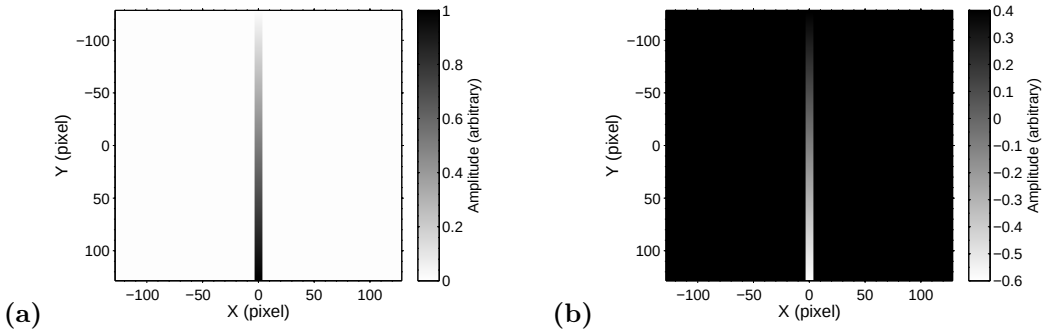


Figure 6.10: Image (a) has a background amplitude $A = 0.01$ and a line seven pixels wide of initial amplitude 0 linearly increasing to 1 along the length. This represents (but does not resemble) a segment of a parabola in a B-scan interfering with a surface wave. Where the line is strong, the largest contribution to the mean is from the line, whereas the median is unaffected by such a thin line. Image (b) has a background amplitude $A = 0.4$ and a line seven pixels wide of initial amplitude 0 linearly decreasing to -1 along the length. This represents either an attenuation, which is only relevant up to where the combined amplitude $A = 0$, or a signal destructively interfering.

to process that scan region for other purposes, and therefore the destructively interfering signal would be lost under those conditions anyway; better the signal be attenuated than lost completely. Usually, in practical cases, at some point, the interfering signal will reach an edge of the horizontal line, and after that point the interfering signal will have a greater amplitude, reducing the severity of the error. Consider if the horizontal line had an amplitude of 1 at each pixel, and the vertical line pixel an amplitude of -8; after subtraction of the median (equal to 1), the vertical line pixel will still have an amplitude of -8 due to the restriction, rather than the ideal case -9. However, the error is only -1 (approximately 10%).

Thresholding introduces artefacts (figure 6.9b), and hence is not used for subtracting horizontal lines.

6.3.2 Reflected surface waves

If the geometry of the situation is known, it can be used to get the angle at which the reflected waves will appear in the B-scan. Other than the common case (figure 6.2), the line will deviate from linear behaviour, but over short distances the deviation is often small enough to approximate

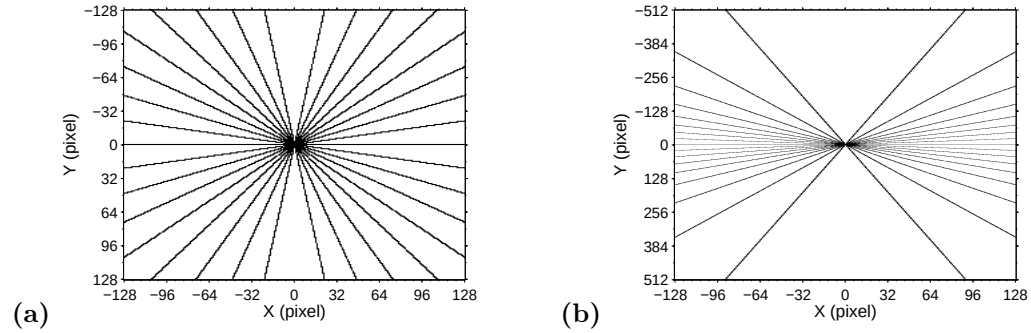


Figure 6.11: These images demonstrate square (1:1) aspect ratio pixels (a) and rectangular (1:4) aspect ratio pixels (b). Although it is clear in this case that the aspect ratio is not square on the right, in the case of a B-scan, a change of sampling rate will change the aspect ratio of the image (not screen) pixels if the image is viewed at a fixed size, without significantly changing the appearance of the image. In that situation, assuming that the actual image display can be assumed approximately square, then it is sensible to use equation 6.23 which will compensate for the change of aspect ratio, and in the case above, would make the lines on the right image resemble those on the left image.

the line as linear. It is important to include mode-converted waves, such as waves that started as a surface skimming longitudinal wave that mode-converted to Rayleigh upon reflection from the edge. The common case and slowest wave mode provide an upper bound to the tested angles.

Detection of lines is performed by the modified HT, with optional grouping such that if there are a large number of detected lines at similar angles, only the angles that are the peaks in a θ histogram are selected. The user selected thresholding parameter, γ, is obtained empirically from trial runs on representative data. If the θ grouping is active, lowering the threshold will only select additional lines if there are local maxima that were previously below the threshold. The IHT is used to draw the selected lines in image space, using the mean value of each line. Where lines overlap, the average value of each line at the intersection is used to assign the point a value. Experimentation demonstrated this to be the safest technique, especially if there are a large number of lines at similar (θ, ρ) values. It has the disadvantage that if there are two genuinely intersecting lines, rather than two lines that are in the same band and intersect due to inaccuracies in the θ estimation, the actual correct value would be the summation of their amplitudes.

It is important to consider the aspect ratio of displayed images (as highlighted by figure 6.11) when choosing which angles to evaluate the Hough transform at, as a linear distribution of angles might not be appropriate. For example, angles that appear to be at 45°, may actually be at a significantly different angle. Consider an image 1000 pixels high and 100 pixels wide. If displayed squarely, the aspect ratio is 1:10 (the pixels are ten times wider than they are high). Consequently, a line appearing at 45° is actually at $\arctan(1/10) = 5.7°$. Assuming a square display to the user, a possible alternative is θ':

$$\tan\theta = \frac{\Delta x}{\Delta y} \tag{6.22}$$

$$\theta' = \arctan\left(\frac{\Delta x_{max}}{\Delta y_{max}} \tan\theta\right) \tag{6.23}$$

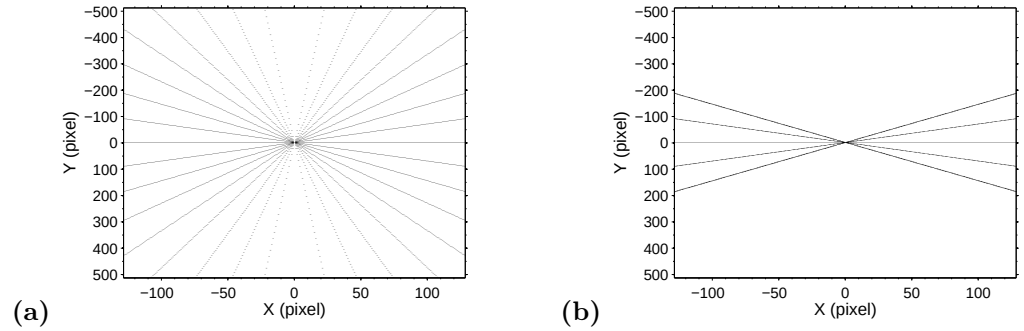


Figure 6.12: Image (a) consists of vertically sparse lines at different angles but at zero offset from the image centre. The modified HT can identify all these lines, and after peak detection, the image can be recovered. The conventional HT can only identify the lines shown in image (b). In both cases, the threshold was $\gamma = 0.2$.

6.4 Results

6.4.1 Simulated

The constraint of no thick lines in the vertical direction is most significant for lines nearest the vertical, and some lines in actual experimental B-scans can appear very sparse at such orientations. Sparse lines created to test the inclusion of the constraint in the HT are shown in figure 6.12a. The modified HT manages to identify all the lines and hence can recreate the original image (after peak identification). The conventional (but using the new threshold criteria) HT fails to identify lines at more vertical angles ($\theta \lesssim 30°$ depending on the exact threshold used). Note that this is a worst case scenario, as lines are often more than a single pixel wide.

Figure 6.12 gives the correct angles to search for and the known offset to the HT algorithm. Whilst this test is illustrative, testing using random angles and offsets is a better measure of the overall algorithm performance. To this end, an automated test program was created that generates a series of 1000 images (255×255 pixels) consisting of nine lines per image, at a continuous range of angles and offsets, $0° < \theta < 90°$, $0 \leq \rho \leq 127$. Lines going in the other direction at the same absolute angle to the vertical or on the other side of the origin have identical detection characteristics and hence are neglected. These images are input into both HT algorithms, and the algorithms check for lines within the same broad range, with $\Delta\theta = 10^{-3}\pi$ radians and $\Delta\rho = 0.5$ pixels. A check is done to ensure that the lines are clearly distinct within each image, since bands are not of interest for this test, but lines are of course allowed to overlap. Noise has not been introduced into the image for this test. The percentage of the total lines that were detected, using the new detection criterion, are summarised in figure 6.13.

At the most vertical angles, the lines consist of only a few pixels for these image sizes, and hence unambiguous identification becomes almost impossible. This is not a major concern, as if it is really is just a few pixels (as for an image this size), it would be impossible to differentiate from noise anyway. For larger images, the number of pixels present increases, and so does the probability of a correct detection.

Lines close to horizontal are expected to lead to very similar performance between the HT algorithms, and this does appear to be the case.

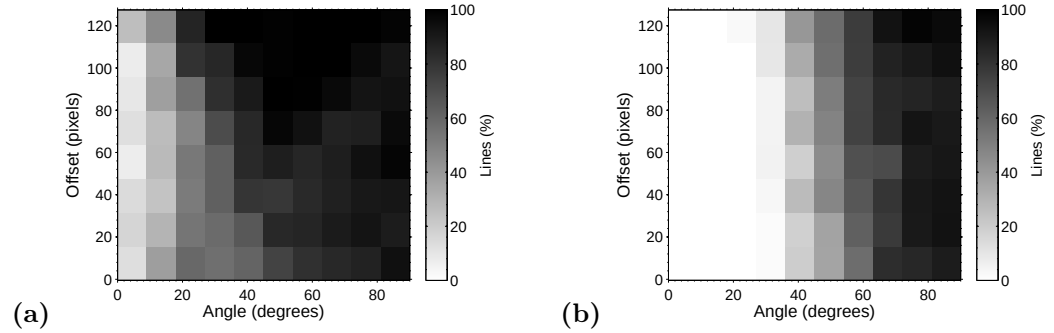


Figure 6.13: HT line detection performance for a random line trial with $\gamma = 0.2$ for scaled (a) and conventional (b) HT algorithms, the scaled HT shows a significant improvement. Performance is better at high ρ as $\Delta\rho$ is much smaller, resulting in more detections but a greater risk of detection duplicates, with certain cases potentially requiring ρ grouping [179].

6.4.2 Experimental

The B-scan in figure 6.14a is formed using a TOFD type setup (figure 1.1a) on a mild steel sample, following the common case shown in figure 6.2; there is also a defect present. Measured wave speeds equalled standard values for mild steel within experimental error, the standard values for compression, shear, and Rayleigh waves being 5960m/s, 3235m/s, and 2996m/s respectively [19]. A Nd:YAG laser, pulsed by Q-switching for 8ns pulses, was focused onto the sample surface and generated ultrasound via ablation. The receiver was an EMAT consisting of a cylindrical permanent NdFeB magnet (N52 grade) linearly wound with 25 turns of 0.08mm diameter copper wire. The magnetic field direction was perpendicular to the sample surface, meaning that the EMAT preferentially detects in-plane motion. On the surface, this allows compression and Rayleigh waves to be detected, and shear waves that are not travelling at an angle very close to the surface. From the bulk, shear waves are strongly detected, and compression waves that are not incident parallel to the field direction are also observed. The characteristics of the emitter and receiver leads to the signals received having some frequency content variation with angle, with higher frequencies being attenuated more strongly along the surface due to the pulses being convoluted with the spatial size of the emitter and receiver. Since for this test the EMAT coil was relatively wide at approximately 2.5mm, high frequencies are strongly attenuated, and the pulses travelling along the surface lose temporal sharpness. A typical pulse duration for a reflected surface wave is approximately 1μs with the majority of the energy being in the region of 100kHz to 1MHz for the surface waves. Despite these examples of pulses not being particularly temporally sharp, the issue of sparse lines is still very significant, and will only get worse if the setup was optimised to generate sharper pulses. Each scan represents a positional increment of 0.8mm and has a sampling frequency of 100MHz. Time-gating has been used to remove the surface-skimming compression wave, which appears too early to interfere with any other waves, and hence need not be considered here.

All B-scans displayed have been band-pass filtered to remove noise components that are outside of the pulse frequency range. Due to the received pulses having frequency content from 0MHz to 3MHz, a lower limit of DC was suitable for the band-pass filter, actually making it a low-pass filter. In this case, the band-pass filter was performed in software after digitisation.

An additional stage of processing was added for the experimental data. The B-scan was split into vertical segments only 10 pixels wide with a 5 pixel (50%) overlap between adjacent segments,

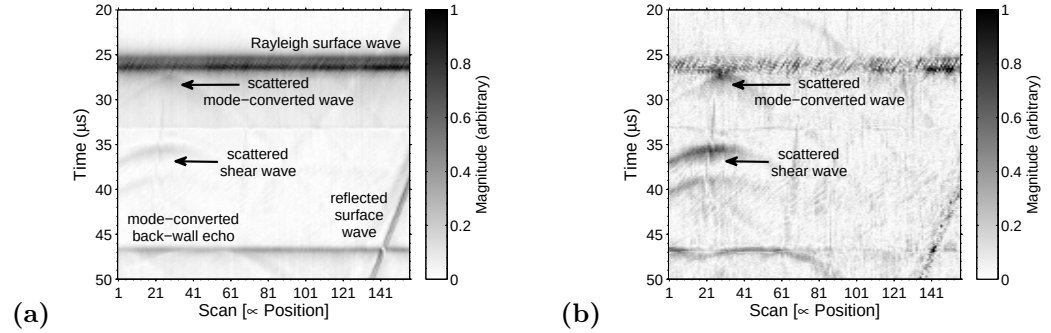


Figure 6.14: The normalised B-scan (a) has strong direct and reflected surface waves in addition to a very weak parabola caused by scattering off a defect. The processed B-scan (b) has had surface waves partially removed.

and each segment was operated on by the HT individually. Lines were found for each segment, and the IHT used to recreate those identified. The segments were then formed back into an entire B-scan, and subtracted from the original B-scan to remove the lines. Splitting the B-scan into segments had two main advantages. Firstly, it deals with the variation of amplitude along the length of a line quite effectively. Secondly, it provides significant computational saving due to the reduced size of the Hough space for each segment. However, it does mean that there are far fewer pixels forming each line, and this could lead to false positives. This was rarely observed in practice, most likely since lines of less than three pixels were ignored. More of an issue was actual lines being missed, and this issue is the subject of future work.

Figure 6.14b shows the B-scan after surface wave removal and a second application of the same frequency filter to remove any high frequency noise introduced by the surface wave removal. Lines have been effectively removed to a large extent. The greyscale axis of the B-scans is an arbitrary linear scale of magnitude, with zero always representing where the magnitude is zero, and one representing where the magnitude is at its maximum. Since the removal of the surface waves reduces the maximum magnitude, weaker waves become darker on the rescaled axis than they were prior to the removal of the surface waves. Before application of the technique, the scattered shear wave only occupied the bottom 20% of the greyscale axis. Now, the scattered shear wave occupies 85% of the greyscale axis. In addition, the scattered mode-converted wave is more visible although it is still partially obscured by the direct surface wave. The reduction of the relative strength of the direct and reflected surface waves allows easier analysis of the data by an operator or by an algorithm such as TOFDI. Unfortunately, the background noise also becomes more visible, but this noise was always present, it was not introduced by the processing.

6.4.3 Reliability mapping

Heavily processed areas may be less reliable for assessing defect positions and sizes due to processing artefacts (generally not significant) or because despite those areas being identified as containing horizontal or angled features, the removal has not been complete. The reliability assessment automatically suppresses such areas (to an operator pre-selected fraction of their amplitude). This can be more aggressively employed if the operator believes it is unlikely that there will be signals due to defects in those regions of the B-scan.

The assessment starts by normalising the data subtracted in the previous stages so that it is

between zero and unity. For this stage, the horizontal and angled lines removed in the previous stage are treated separately. The data is then multiplied by a user selected value between zero and unity, so that when it is used in the reliability mapping, such areas are only suppressed as much as the user considers necessary. Then, again treating the horizontal and angled lines separately, the data is stretched in time by convolving it with a Hann window that has a user selected length, and the absolute magnitude is taken. The result is then subtracted from one and multiplied with the signal data, so that values without data that needs suppressing will be multiplied by unity, whereas areas that do need suppressing will be multiplied by a number less than unity down to a limit of zero.

Lines not detected in the previous stages are also suppressed (rather than removed, which is more difficult using the second, supplementary method). The secondary line detection makes a copy of the B-scan, applies a median filter (an extreme form of outlier rejection [168]) along the scan direction, and takes the magnitude of the analytic signal along the time dimension. This magnitude data is copied and will be used later. Differentials are taken along the scan direction of the magnitude as angled lines appear rapidly in the scan direction, whereas parabolas appear gradually (both appear at the same rate along the time axis, hence differentials are not taken for that dimension). Identified points are stretched along the time dimension to ensure they include most of the pulse. This stretching is done using a bipolar matched filter of the length of the pulse width convolved, similar to what is done for the horizontal and angled lines; the absolute magnitude is then taken. The minimum value is then subtracted and the result normalised so that it extends from zero to one. The result of this is multiplied by the same user selected value between zero and one that was used for the horizontal and angled line data. As differentials were taken, the reliability map for this stage is one scan short of the data to which it will be applied. To rectify this, the map is copied and shifted by one scan column. The maximum value from the original map and the shifted map for any given column is taken, which can have the effect of stretching large amplitude suppression points to two scan positions, a generally quite useful side-effect as it makes it more likely that large amplitude points that should be suppressed in the signal, will be suppressed. This current reliability map is multiplied by the magnitude data that was copied earlier, and normalised again so that it is between zero and unity; this means that higher amplitude identified points are suppressed much more than lower amplitude identified points. This is necessary as positions that show up strongly in the identification stage are much more likely to require suppression than points that showed up weakly, and it minimises the impact on points that have been incorrectly (but only weakly) identified. As with the other stages, the result is subtracted from one and multiplied with the signal data, so that values without data that needs suppressing will be multiplied by unity, whereas areas that do need suppressing will be multiplied by a number less than unity down to a limit of zero.

In practice, the reliability assessments are combined and applied to the B-scan in one go. The result of the combined reliability mapping can be seen in figure 6.15.

This reliability mapping is primarily for the benefit of a operator viewing the B-scans before they have been processed into a cross-sectional image; it is not generally used when converting the B-scans into a cross-sectional image. One of the purposes of removing the direct and reflected surface waves, is that it allows any scattered waves that have been obscured to be observed and used in the cross-sectional imaging. By suppressing these regions, these regions cannot be correctly used in the imaging process, but an operator examining the B-scan can use the suppression to improve the dynamic range of the B-scan if necessary. If the user selects the maximum possible suppression,

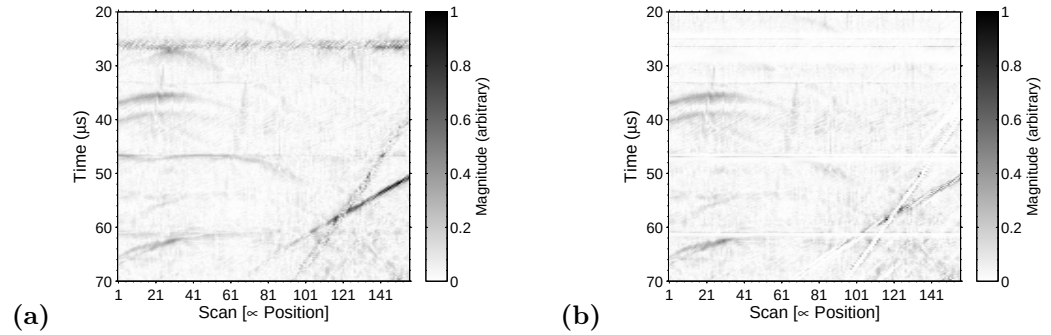


Figure 6.15: A B-scan (a) has already had horizontal and angled lines removed using the methods described in section 6. Affected regions can be marked as unreliable and suppressed if necessary, and features that change sharply along the scan dimension but have so far been missed can be identified and also suppressed (b). The purpose of this reliability mapping is to minimise strong features that are not signals scattered from defects, as such features will hamper the imaging stage. The magnitude axis of the B-scans has in both cases been scaled to go from zero to the maximum magnitude present in that B-scan.

the B-scan can then be used in the cross-sectional imaging process, as the reliability mapping is effectively operating as a more complicated form of traditional time-gating, but again, this is not usually used.

6.5 Future work

Ignoring attenuation, omnidirectional surface wave amplitude is proportional to $1/\sqrt{r}$ which can often, but not always, be neglected. In the current work, this has been handled by segmenting the image into vertical slices, but this reduces the information available to detect the line. Work is being done to combine multiple slices to use all the information whilst accounting for changes of line amplitude.

Smaller peaks are sometimes only detected after the larger peaks have been removed successfully. Multiple passes of the HT has been tried but was not found to be particularly successful. An alternative is to model what a specific line in image space appears as in Hough space, and then for each line removed from the image space, to remove the corresponding shape in the Hough space [18, 173]. Another option is to HT just that one identified line and subtract that from the Hough space, a less elegant but computationally cheap solution. These techniques are still being explored and no definitive conclusion of which is most suitable has been reached at this stage.

Alternative parameterisations [178] must be considered, as the accuracy of the (θ, ρ) parameterisation decreases as ρ increases. For example, the "Muff" transform [180] parameterises a line by the points at which the line enters and leaves the image and there is a variation which uses an entry point and an angle. Care must be taken when modifying the parameterisation to suit the constraints outlined in this work.

Two avenues that are not of interest are the inverse voting algorithm for the HT [174], which is unlikely to provide much benefit here as the HT is only 2D anyway, and the probabilistic Hough transform [33, 181]. The concept of the latter is that if just a small random subset of the points in the image are used as input to the Hough transform, the performance of the transform is still satisfactory. However, prior work using this technique [181] uses lines that cross the longest part

of the image when analysing performance (corresponding to the part of the Hough space with the most votes for a uniform image). Lines analysed in B-scans might populate part of the Hough space that has far fewer votes from a uniform image, and hence the performance will drop off much faster due to the use of a subset of the input, and the transform will become much less robust to noise for these cases. In addition, there may be a large number of lines at similar ρ, and this may not be amenable to the probabilistic Hough transform.

6.6 Conclusions

TOFD B-scans may contain strong features, due to surface waves propagating directly from emitter to receiver, bulk waves reflecting off a back-wall, and surface waves reflecting off a surface edge, which can mask bulk waves scattered by features of interest within the sample under test. Reflected surface waves can form lines in the B-scan at angles less than 45° from the time axis, which will contain gaps and are considered sparse unless the ultrasound pulse is temporally very broad. The problem is worse for temporally sharper signals or coarser scanning movement, and is still a problem for relatively small scanning increments of 0.8mm and surface waves that at 1μs are relatively long. Such B-scans are produced by the laser-EMAT system used to collect the experimental data shown in this work.

It has been demonstrated that direct surface waves can be dealt with very simply by subtracting the median along the scan axis at each time point, and clamping the result such that the magnitude of any specific point cannot be larger than its previous value, although the sign of the amplitude can change. A standard HT is not designed to deal with sparse lines, but the modified HT incorporates the constraint that a sparse line can exist along the time dimension by rescaling the B-scan, and it has been demonstrated that this successfully removes reflected surface waves from B-scan images generated as part of the TOFD(I) technique. The removal of strong surface waves, direct and reflected, from the B-scan, reveals the scattered bulk waves and can significantly improve manual or automated scan analysis.

This technique could potentially be applied to other processes that can form a B-scan. Such applications are currently speculative, but an example has been included here for consideration. If a 2D linear array operating in FMC mode is positioned at the edge of a sample, reflections of surface waves from the edge will be observed. In this case, the array does not need to scan to observe the problem, as the elements at different positions will produce the same effect. If the data from each receiver (considering a single emitter) was formed into a B-scan, the wave reflected from the edge could be identified and removed using the modified Hough transform technique. This could be repeated for each emitter in the array (each one forming its own B-scan), and the reduction of artefacts could improve imaging performance at the sample edge.

Finally, this technique could also be applied to images that are not sampled equally along x and y directions.

6.7 Use in development of TOFDI

This chapter has enabled the removal of a considerable amount of coherent interference from ultrasonic B-scans. TOFDI can output processed B-scans and cross-sectional images, and this work helps improve both of these. B-scans gain dynamic range, making scattered waves more

visible, and removal of irrelevant straight lines makes interpretation of the B-scans easier. Cross-sectional images can be generated without the coherent interference creating problems, resulting in fewer artefacts in the final output. Again, dynamic range is improved (since very high magnitude straight lines can produce very large and bright artefacts), and so the cross-section becomes more interpretable, as there are fewer false defect indications.

Chapter 7

Parabola filter

This chapter has been published [182] in a modified form in a peer-reviewed journal.

This chapter describes a method for filtering B-scans that reduces noise, and transforms parabolas (caused by ultrasonic wave scattering by defects) to points, using a parabola matched filter. This is an alternative to the cross-sectional imaging process described in chapter 8, and was created to convert a B-scan into something resembling a cross-sectional image, but still retaining the standard B-scan time and scan position axes, although these can be converted to depth and lateral defect position respectively. Multiple B-scans are also combined into a single B-scan where possible, allowing for the merging of multiple data sets. The method is fast, only using a relatively small number of FFT operations, and easy to implement. Both the filtering and the combining make visualisation of the B-scans output from TOFDI a simpler task, with an operator able to use the filtered B-scan to identify and position defects, whilst having the original data available, in the same format, for confirmation.

7.1 Introduction

The change in time of arrival of the diffracted ultrasound pulse, as a B-scan moves over a defect (figure 7.1), forms a characteristic parabola shape which human vision can clearly distinguish, and this can be analysed to determine defect position and size. Unfortunately, manually distinguishing a parabola is expensive in terms of time and operator training, and hence it is desirable to make the process easier or even to automate it.

The HT [166], which can detect parameterisable shapes [167, 168], has been used in prior work to locate parabolas in B-scans [17, 18, 34, 183] using an approximation of the shape [17]. These techniques often use some method to improve computational speed, such as an IHT [174] and genetic algorithms to solve the consequent parameter optimisation problem. However, these procedures are relatively complex to implement, and a simpler technique, such as the parabola matched filter described here, could be an attractive alternative.

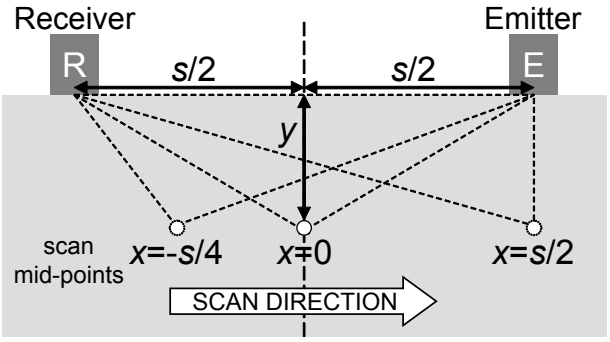


Figure 7.1: The diagram depicts the classic TOFD setup, with a defect at depth y and lateral position $x = 0$. The emitter-receiver pair scan across the sample with a constant separation s between transducers, using a constant scan motion parallel to the line between the emitter and receiver. The emitter-receiver mid-point moves from $x = -s/2$, through $x = 0$, to $x = s/2$. The scattered shear wave parabola in the B-scan is formed as the emitter and receiver scan over the defect, with the peak when the mid-point is at offset $x = 0$. The distance increment is sufficiently fine for any given parabola to exist over a reasonable number of scans.

7.2 Theory and method

7.2.1 Parabolas in B-scans

The motion of the emitter and receiver relative to the scatterer describes the characteristic parabolic shape. Consider a scattering point at a depth, y, and at a lateral position, $x = 0$, as shown in figure 7.1. The transducers are separated on the surface by a distance, s, and are scanned in a direction parallel to the line between the transducers. For the purposes of parabola matched filtering, it is assumed that the scattering point lies on the plane containing the transducers, perpendicular to the surface. Then, for an emitted wave propagating at speed c_1, and a scattered wave propagating at speed c_2, the propagation time, t, as the transducers scan across the surface, is given by simple geometry:

$$t(x,y) = \frac{\sqrt{(s/2-x)^2+y^2}}{c_1} + \frac{\sqrt{(s/2+x)^2+y^2}}{c_2} \tag{7.1}$$

Often the parabola can only really be seen in the spatial range $-s/2 < x < s/2$ since outside of that range, transducer directivity and other factors can cause the received amplitude to drop below the sensitivity threshold.

7.2.2 Parabola matched filtering in B-scans

A matched filter will detect the signal to which it is matched with the optimum signal-to-noise ratio when the noise can be considered white [146], and can be applied to the received signal using cross-correlation. Practically, this is usually performed on discrete data using a FFT [156] due to the considerable speed benefits. In this case, the signal is not a one-dimensional signal, but a two-dimensional parabola in a B-scan, defined by equation 7.1. For a given B-scan, s is fixed, c_1 and c_2 are assumed to be only able to take the values of the bulk compression or shear waves, and by varying x, a parabola filter can be described for a chosen y. Using cross-correlation, this parabola matched filter can simply and quickly be applied to the B-scan, and this will detect parabolas generated by point-like scatterers. However, the form of the parabola will vary with the depth of

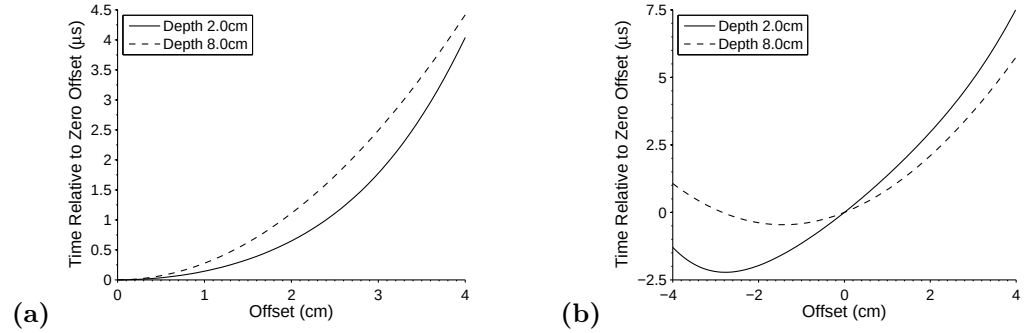


Figure 7.2: Parabola shapes for defects at depths, y, of 2cm and 8cm, are shown for the $c_1 = c_2 =$ 3226m/s case (a) which is symmetric, and the $c_1 \neq c_2$ case (b) with $c_1 = 3226$m/s and $c_2 = 5932$m/s which is asymmetric (the speeds are experimentally measured shear and compression wave speeds respectively, for a mild steel sample). The separation between the emitter and receiver, s, is 8cm. The zero offset time for each parabola has been subtracted from the parabola to form $\Delta t(x, y)$ as defined by equation 7.2. As the offset, x, is varied, $\Delta t(x, 8\text{cm}) - \Delta t(x, 2\text{cm})$ is always positive for (a) as long as $-s/2 \leq x \leq s/2$, but varies between positive and negative values for (b).

the scatterer, y, as can be seen in figure 7.2. Therefore, the parabola filter can only be considered matched for the particular depth it was created for. To match perfectly for a large depth range of closely spaced depths, would require application of a large number of matched filters to the B-scan, which would be prohibitively computationally expensive. If a large increment is used for the change in depth between parabola filters, it is possible that the filter will hinder rather than help analysis due to the potentially poor correlation between the signal in the B-scan and any of the matched filters applied. This work describes how to cover the depth range of interest, whilst minimising the number of required matched filters, and keeping the difference between the matched filter and the actual parabola within bounds that retains the effectiveness of the matched filter.

Approximations to the shape of the parabola have been made by other workers [17], but this is not a requirement in this case, as the full description of the parabola can be used. The implementation of the parabola matched filter described here relies on the ultrasound signal having a finite and reasonably constant bandwidth, such that the pulse amplitude remains at the same polarity (either positive or negative) for a known minimum time, T, that is consistent along the length of a given parabola (but not necessarily consistent between separate parabolas). As an example, consider a signal which along the time dimension consists of a single cycle of a sine wave, of period $2T$, as in figure 7.3a, which has also had a customised raised cosine window applied. The pulse has negative amplitude for time T and then positive amplitude for time T. For a perfect match, the parabola matched filter would have the exact same shape as the parabola in the B-scan, but it is sufficient in many cases if the filter parabola and the actual parabola are merely the same sign along the length of the parabola. If the matched filter is an infinitesimally thin line, as long as it does not deviate from the parabola formed by the sine wave by more than $T/2$ in either the positive or negative time direction, positive parts of the sine wave are always contributing to the point that the matched filter will form, not the negative parts (or vice versa). This means the filter can stray from the exact shape, in either direction, by a maximum of half the time for which the amplitude does not change sign (from positive to negative or vice versa). The result is that a single parabola acting as a matched filter can act upon a range of parabolas at different depths in the B-scan whilst maintaining effectiveness (figure 7.3b), as the matched filter is sufficiently close to the same shape as the actual parabola that it only includes the part of the parabola that is

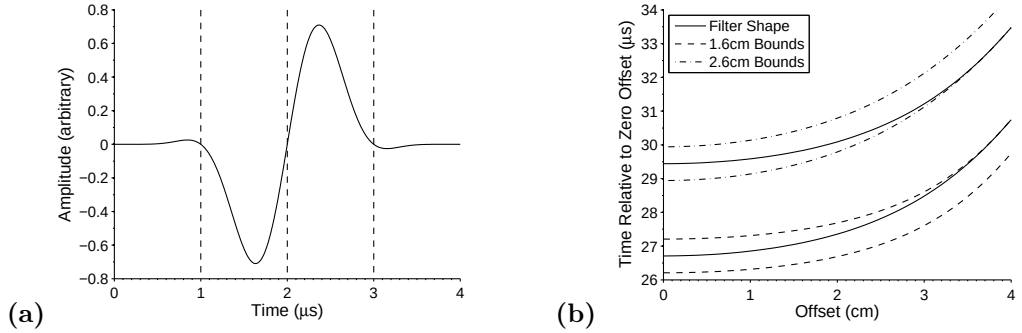


Figure 7.3: Consider an example simulated signal of centre frequency 500kHz (a); the signal is negative for 1μs (the region between the first and second dashed lines) and then positive for 1μs (the region between the second and third dashed lines), and therefore $T = 1\mu$s. A single prototype parabola can be used to filter a range of depths if the signal has a period of this length. Assuming that the received parabolas observed will have a temporal width of 1μs, for the symmetric $c_1 = c_2 = 3226$m/s case with separation 8cm, bounds are shown (dashed and dot-dashed lines) at $\pm T/2$ from the parabolas for scatterers at $x = 0$cm with depths of 1.6cm and 2.6cm (b). The prototype parabola (exactly matched for a depth of 2cm) is cross-correlated with the B-scan in both temporal and spatial dimensions, and it can be seen that it would sit within either set of bounds if appropriately shifted in the temporal domain (solid lines). The matched filter would respond to both the positive and negative regions of the signal.

all one polarity (either positive or negative), not a mix of the two. A relatively small number of parabola filters can then cover the entire depth range of interest.

Returning to the example signal of a single cycle of a sine wave, it is still the case that the closer the matched filter to the parabola of interest, the stronger the filter response will be, as if the two have the exact same shape, then all the highest amplitude parts of the sine wave (close to 1.6μs or 2.4μs in figure 7.3a) will contribute to the peak of the signal after matched filtering, making that peak larger. With cross-correlation, the response seen to a perfect match will be the auto-correlation of the matched filter, but in this case the matched filter is a infinitesimally thin line, whereas the actual parabola must have a finite width in the temporal direction, and this means the auto-correlation of the matched filter is not particularly relevant. In practice, rather than actually being infinitesimally thin, the filter is one sampling period wide, and if the actual parabola were also this wide (if the pulse was temporally very sharp), then the matched filter and the parabola shape would have to match exactly, and the response would nearly be a Dirac delta function in the two-dimensional B-scan. This is a worst case scenario for the technique, as $T \to 0$ and the number of required parabola filters tends to infinity. The technique works much better if the signal is closer to a rectangular function, as then the response is the same as long as the filter shape does not move beyond the bounds of the parabola formed from the rectangular function for small deviations from the correct depth.

The filters can be applied for each c_1 and c_2 combination of relevance, and the results combined.

7.2.3 Implementing parabola matched filtering

This section describes the procedure for using the parabola matched filtering technique. The positional increment of the B-scan is assumed constant.

Values are selected for c_1 and c_2, with mode-conversion being represented by $c_1 \neq c_2$ (in which case the parabola is asymmetric about $x = 0$). If the emitter is assumed to be the $(s/2 - x)$ side of

equation 7.1 (on the "right", as in figure 7.1), then c_1 represents the speed of the generated wave and c_2 represents the speed of the wave after scattering off the defect; this is reversed if the emitter is assumed to be the $(s/2 + x)$ side of equation 7.1 (on the "left").

Usually it is necessary to explore a range of depths. To filter rapidly, it is beneficial to use as few different parabola filters as possible, otherwise a region will have to be filtered a large number of times since a parabola filter extends in time and across the scans. Each parabola exists over a different temporal range, with the spatial (but not temporal) centre of the parabola being at $x = 0$, where the emitter and receiver are the same distance from the scatterer, and the extrema of the parabola being at x_{max} or x_{min}. The time Δt represents the temporal range of the parabola relative to the centre:

$$\Delta t(x, y) = t(x, y) - t(0, y) \tag{7.2}$$

This allows parabolas at different depths to be compared in terms of shape without consideration of the arrival time for their spatial centre, as in figure 7.2. It is computationally quite cheap to evaluate equation 7.1 for a range of different y values, and then to subtract $t(0, y)$ to form $\Delta t(x, y)$ for each y value. Next, starting at the initial (most shallow) depth of interest, y_1, $\Delta t(x, y_1)$ is compared against $\Delta t(x, y)$ at every other deeper y value, with a suitably small increment, until either $|\Delta t(x, y) - \Delta t(x, y_1)| > T/2$ at any x value (when a prototype parabola filter for the initial depth of interest y_1 would have a shape that would move it outside the single sign region of the parabola signal at depth y), or y_{max} is reached. From the evaluated depths, the last depth at which $|\Delta t(x, y) - \Delta t(x, y_1)| < T/2$, over all x values, is then selected as the first prototype filter depth, $y_{f1} = y_2$. The last depth is chosen, because if a parabola for $y_1$could filter a parabola at y_2, then the inverse is also true. There is an advantage in picking y_2, as it can also filter everything within the range $|\Delta t(x, y) - \Delta t(x, y_2)| < T/2$, a limit referred to as y_3. This process can then be repeated; the last depth at which $|\Delta t(x, y) - \Delta t(x, y_3)| < T/2$, over all x values, is then selected as the second prototype filter depth, $y_{f2} = y_4$. And therefore the process repeats, prototype filters being at y_2, y_4, y_6, and continuing to $y_{fn} = y_{2n}$ where $n \in \mathbb{Z}^+$, with each filter representing the range $y_{2n-1} \leq y_{2n} \leq y_{2n+1}$, until y_{max} is obtained as an upper limit. As a depth increment, $\Delta y = \sqrt{(c/2f_s + s/2)^2 - (s/2)^2}$ with sampling frequency f_s should be sufficient, although in practice the range of depths does not even need to increase linearly. The x value at which the parabolas are most different has no exact analytical expression that can be formed from $\frac{\partial}{\partial x}|\Delta t(x, y) - \Delta t(x, y_1)| = 0$, but comparing points numerically takes negligible time on modern hardware (and certainly is not the bottleneck of the parabola matched filter), the result of which can be seen in figure 7.4. The comparison is performed at points with the same x values for both parabolas being compared (but for all x offsets), where $0 \leq x \leq x_{max}$ for symmetric parabolas representing scattering without mode conversion and $-x_{max} \leq x \leq x_{max}$ for asymmetric parabolas representing scattering with mode conversion. The user selected parameter x_{max} is typically $x_{max} = s/2$. The direction of $\Delta t(x, y) - \Delta t(x, y_1)$ could be positive at one x value and negative at another, as can be seen in figure 7.2 for the mode-converted case. For the standard case without mode-conversion, $[\Delta t(x, y) - \Delta t(x, y_1)] \geq 0$ for $-s/2 \leq x \leq s/2$. Using the magnitude, $|\Delta t(x, y) - \Delta t(x, y_1)|$, greatly simplifies the parabola comparison as the change could be in either direction, and using $T/2$ rather than T means that it can be assumed that the parabola filter and the actual parabola will at least match well close to the $x = 0$ point. Bounds of $\pm T/2$ from parabolas at different depths, and how a prototype parabola from between the two

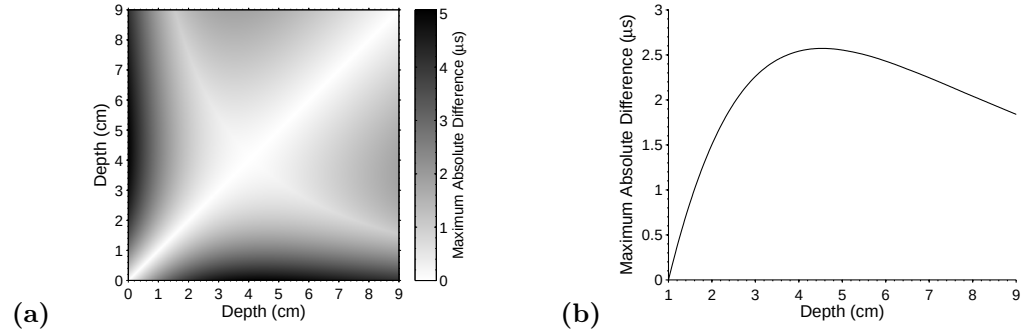


Figure 7.4: The maximum absolute shear wave parabola differences, for $c_1 = c_2 = 3226$m/s and an 8cm transducer separation (a), can be used to evaluate if a new parabola matched filter is required when moving from one depth to another. Depths form both the axes, y_1 and y_2, and the greyscale represents the maximum $|\Delta t(x, y_2) - \Delta t(x, y_1)|$ for each (y_1, y_2) pair over all x values. Comparing a depth of 1cm with all other depths, again taking the maximum difference over all x values, the difference is clearly non-linear with respect to the change of depth (b). The x value at which the maximum occurs varies with depth. Note that (a) is symmetric about the diagonal from $(0, 0)$, to $(9, 9)$, and thus only half of the differences need be evaluated or stored.

depths fits within those bounds, is shown in figure 7.3b; parabolas between those two depths could be filtered with the same matched filter if formed from a signal such as that shown in 7.3a.

Each selected y_f is used to create a parabola matched filter, using the depth as an input into Δt and the values of Δt at each x value to mark (set equal to one) the appropriate time element in an initially clear image. The image acting as a matched filter is then a line forming a suitable parabola, with a temporal width of one sampling point at any given offset. The zero time and spatial offset point of Δt forms the first element of the image, and half the filter (r points, the filter being of length $2r + 1$ in the time dimension) appears in the negative spatial side of the filter, analogous to the position of negative frequencies in FFT output. Similarly, negative time parts of the filter appear in the negative temporal side of the filter. Having the zero offset point of the parabola at $(0, 0)$ in the image ensures that the response to a matching parabola appears at the correct depth and position in the final B-scan output. Figure 7.5a gives part of the positive spatial side of an example parabola matched filter, and figure 7.5b is the magnitude of the parabola to which it is matched.

How many parabolas are required to filter a specific depth range is dependant on factors such as the separation of the transducers (s), the exact depth (y) range of interest, the speed of the waves (c), the time for which a received pulse remains at the same polarity (T), and over what scanning range the parabola exists (x). The number of parabolas required for any given set of parameters is easy to calculate numerically. It is not possible to provide a general answer to the question of how many parabolas are required, due to the parameters involved, so only specific cases can be considered. Assume that $-s/2 \leq x \leq s/2$, the pulse polarity remains the same for 2µs, and that the depth range is 0cm to 10cm. For a transducer separation of 8cm, and a wave speed of 6000m/s (used simply as an approximate value for compression waves), the depth range can be covered by only two parabolas, one for depth 0.6cm, and one for any depth between 1.5cm and 10cm (the y_{max}). If the separation is 4cm, only a single parabola is required, at a depth of 0.8cm. If the separation is 12cm, three parabolas are required, at 0.6cm, 2.3cm, and any depth between 2.9cm and 10cm (the y_{max}). Switching to a speed of 3000m/s (an approximate value for shear waves), a separation of 8cm requires three parabolas, for depths 0.3cm, 1cm, and 2.2cm. A separation of

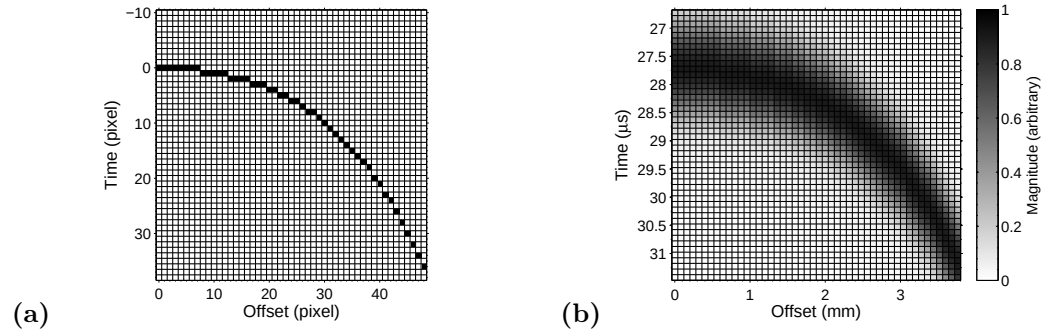


Figure 7.5: The positive spatial side of an example parabola matched filter for a separation of 8cm, a depth of 2cm, and a speed of 3226m/s, is created by using the depth as an input into Δt and the values of Δt at each x value to mark (set equal to one) the appropriate time element in an initially clear image (a). Each spatial element represents 0.8mm, and each time element represents 0.1µs in this example. Gaps along the time axis occur if the change in time is greater than a single sampling period for a single change in offset position; this is the correct behaviour. Negative time pixels, of which there are none in this example as it is not a mode-converted parabola, would actually appear at the bottom of the filter (off the scale in this case). Space is only given to them here so that the matching of the filter and the equivalent signal is clear. The equivalent signal that would be seen in a B-scan has the same shape but is temporally wider (b). For clarity, the magnitude is shown rather than the amplitude.

4cm requires three parabolas, for depths of 0.3cm, 2cm, and a depth between 4.5cm and 10cm (the y_{max}), and a separation of 12cm requires five parabolas, for depths of 0.3cm, 0.9cm, 1.7cm, 2.9cm, and a depth between 3.9cm and 10cm (the y_{max}). Even if it were possible to answer how many parabolas are required in general, rather than for specific cases, the question is not necessarily important, as the speed of the parabola filtering operation is also dependent on the range each parabola filter covers, not just how many there are.

7.2.3.1 Applying the filter to the B-scan

For a two dimensional image, in this case a B-scan, $p(m, n)$, of size (M, N), the DFT, $P(k, l)$, for any specific component, k for spatial frequency components along the first dimension (previously the B-scan position axis, m), and l for frequency components along the second dimension (previously the B-scan time axis, n) is obtained by:

$$P(k,l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} p(m,n) e^{-j2\pi km/M} e^{-j2\pi ln/N} \tag{7.3}$$

Assuming $(k = K - 1, l = L - 1)$, the inverse two dimensional DFT is then given by:

$$p(m,n) = (1/KL) \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} P(k,l) e^{j2\pi km/K} e^{j2\pi ln/L} \tag{7.4}$$

Note that $0 \leq m \leq M-1$, $0 \leq n \leq N-1$, and $0 \leq k \leq K-1$, $0 \leq l \leq L-1$. Usually $M = K$ and $N = L$, but if zero padding is used after the forward DFT is performed, this is not the case, in which case $M \leq K$ and $N \leq L$. In this case, there is no zero padding after the forward DFT, only before it. The cross-correlation in the original B-scan domain, $C(m, n)$, is given for a B-scan $p(m, n)$ and matched filter $p_0(m, n)$ with DFT space representations $P(k, l)$ and $P_0(k, l)$ respectively, by:

$$C(x,y) = (1/MN) \sum_{l=0}^{M-1} \sum_{k=0}^{N-1} P(k,l) P_0^*(k,l) e^{j2\pi km/M} e^{j2\pi ln/N} \tag{7.5}$$

If the FFT is used instead of the DFT, this becomes the standard FFT cross-correlation method [156] in two dimensions.

Since a given filter is only intended to cover a specific depth range, it is not necessary to apply every filter to the entire data range; consequently the B-scan temporal range $t_1(0,y_1) \leq t \leq t_2(x_{max},y_3)$ is chosen, as this completely encompasses the filter range but no more. When using the FFT to perform cross-correlation, to ensure linear correlation over the data range of interest, the standard technique of zero-padding is used. The size of the zero-padding is determined by the maximum size (not including the zero point) of the filter in either the positive or negative direction. As implemented here, half the filter is positive spatial offsets and half negative spatial offsets, and only half the size (minus the zero offset middle point) must be added in zero-padding the spatial dimension, to a size of $\frac{3}{2}(m-1)+1$. For parabolas that are not the result of mode-conversion, the entire filter is in the positive side temporally (excluding the zero time first point), and hence the filter must be zero-padded to twice its size (excluding the zero offset middle point), to a size of $2n-1$ in that dimension. For mode-converted parabolas, the larger of the positive or negative temporal extents is used as the value n before zero-padding to a size of $2n-1$. The data is then zero-padded to a size equal to the nearest power of two (equal to or greater than the current size) in each dimension, although in practice, the two zero-padding operations are combined. Power of two zero-padding is advised since most FFT implementations are fastest for data of size 2^a, $a \in \mathbb{Z}^+$, and the filter is to be applied via FFT. The filter is zero-padded to the same size as the data (enlarging the middle section, between the positive and negative spatial sides), as this is a requirement for an FFT applied cross-correlation to be performed. The filter is applied by multiplying together the two-dimensional FFT of the selected signal data and the complex conjugate of the two-dimensional FFT of the filter.

After the filter is returned to the original spatial-temporal domain by performing a two-dimensional IFFT, the zero-padding is removed in both dimensions (since the zero-padded regions only contain cyclically cross-correlated data of no interest), and the data is trimmed to $t_1(0,y_1) \leq t \leq t_2(0,y_3)$, the data range over which the filter was valid; the extra part $t_2(0,y_3) < t \leq t_2(x_{max},y_3)$ was so that the filter operation included the tails of the filter for parabolas starting at $t_2(0,y_3)$, but the filter is not valid for parabolas starting in the range $t_2(0,y_3) < t \leq t_2(x_{max},y_3)$ and hence this data can be discarded after it has been used in the filtering operation.

The filtered data is stored and the process repeated for the next range, y_3 to y_5 (where y_3 replaces y_1, and y_5 replaces y_3), until all the depths of interest (up to y_{max}) have been filtered. In many cases, for pulses of frequency approximately 1MHz, only a few different filters are required, most of them at the smaller depths where the curvature changes most rapidly (figure 7.4), making the process efficient.

The final stage is to replace data in the original scan with the filtered data for the range $t_1(0,y_1) \leq t \leq t_2(0,y_{max})$, leaving regions outside of the filtered region as the original data. This stage is purely for aesthetic purposes, the original data can be discarded in the regions outside the filtered range if desired. In some cases, the regions outside those filtered may only contain information that cannot be related to a single scattering event from a bulk defect (i.e. before the direct surface wave for that wave speed, or after the back-wall echo for that wave speed).

To work well, it is expected that the parabola amplitude in the B-scan that is being filtered

drops off sharply as it moves away from the correct depth in the cross-correlation, but not to the extent that $T/2$ is so small that a large number of different filters are required. The filtering is repeated for each wave speed and mode conversion of interest, and then the results are combined. If only the filtered data is required and not data outside that region that has not been filtered, then the B-scans can be combined by scaling each time axis such that it becomes a depth axis, using linear interpolation and a rearrangement of equation 7.1 with $x = 0$ (the centre of the matched filters):

$$y(t) = \frac{\sqrt{4c_1^2c_2^2t^2 - s^2(c_1 + c_2)^2}}{2(c_1 + c_2)} \tag{7.6}$$

It is a simple matter to take the mean of the output of each filter where they overlap. In addition, the scan number axis is converted to a lateral position axis, with each scan being assigned the same position as the point half-way between the emitter and receiver, the $x = 0$ position, for that scan.

7.3 Results

Parabola matched filtering has been tested within TOFDI [142, 143, 163]. The experimental B-scans (figure 7.7 and figure 7.8b) are generated using a TOFD type setup (figure 7.1a), using a pulsed laser to generate ultrasonic waves via ablation [6] and a linearly wound EMAT [89] to receive them. Although this setup can generate and receive multiple wave modes, including compression and shear waves and conversions between the two, only shear waves (without mode conversion) are present in the B-scan shown here (figure 7.7a). Prior to parabola matched filtering, the B-scans, as part of the TOFDI process [163], are band-pass filtered, and have horizontal and angled linear features (direct and reflected surface waves respectively) automatically removed from the image [143]. For B-scan display purposes, the reliability mapping has been used with selected areas reduced to 10% of the value they had before mapping (as discussed in chapter 6), but this was not necessary when performing the parabola filtering. The B-scan simulation (figure 7.6 and figure 7.8a) is only designed to mimic real data using very simple ray casting, for the purposes of testing signal processing algorithms. It uses equation 7.1 to produce the parabolas. As in the experimental case, only shear waves (without mode conversion) are present in the B-scan here (figure 7.6a). For both simulated and experimental cases, $s = 7.8$cm, $c = 3226$m/s, and $T = 1\mu$s. B-scan images before and after filtering are displayed with magnitude as the luminance for clarity, but it is instantaneous amplitude that is the input and output of the parabola matched filter. The greyscale axis is scaled for each figure such that zero on the scale always represents zero in the data, and one on the scale represents the highest magnitude point for that data set.

Considering the experimental data, the relatively large temporal width leads to a temporally and spatially wide response to the filter. Although the initial SNR is actually quite good, the short length of the parabola restricts the ability of the filter to reduce random noise optimally and increases the width of the filter response still further. Despite this, the peak of the filtered response has a value over 50% higher than the peak of the unfiltered parabola relative to the background noise, and examining the region around the peak of the filtered response, the SNR is approximately twice that of the region around the peak of the unfiltered parabola. The simulated parabolas have the same temporal width, but are much longer spatially, so they respond much better to the filtering, despite the presence of strong AWGN. For either data set, the peak of the response should still be at the correct depth and scan position. Testing against simulated mode-converted

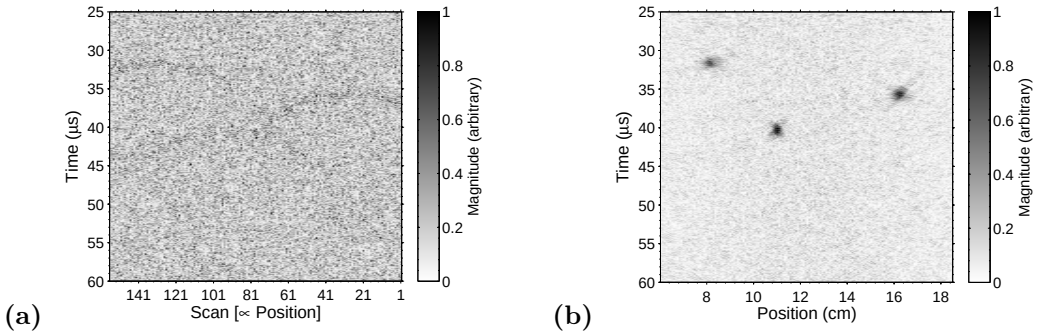


Figure 7.6: Simulated B-scans, before (a) and after (b) parabola matched filtering. The transducers are 7.8cm apart and the scan is in 0.8mm increments; only shear waves are emitted and received. Three scatterers have been simulated, leading to three parabolas (a) and three peaks (b). AWGN has been added to the simulated B-scan (a), with a standard deviation half the maximum amplitude of the parabolas. This makes the parabolas difficult to see with the eye, but the case is still close to ideal for the parabola matched filter due to the long and perfectly curved simulated parabolas, hence the peaks appear clearly in the correct scatterer positions (b).

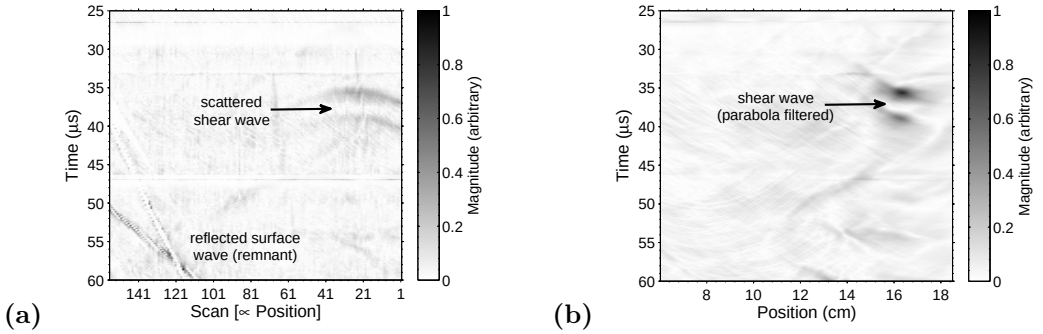


Figure 7.7: Experimentally obtained B-scans, before (a) and after (b) parabola matched filtering. Prior to this, the B-scan has already been band-pass filtered, and had direct and reflected surface waves removed automatically (although the removal is not complete, as can be seen from the remnants in the scan). The laser and EMAT are 7.8cm apart and the scan is in 0.8mm increments. Two parabolas appear in the B-scan (a) due to shear waves scattering (without mode-conversion) from the top and bottom of the defect, hence the two peaks in the filtered data (b).

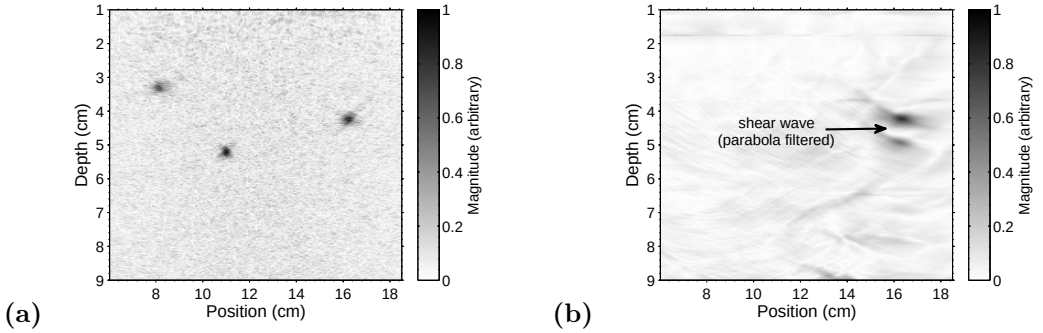


Figure 7.8: Corresponding to figure 7.6b (a) and figure 7.7b (b), these B-scans have been re-scaled (using interpolation) such that the vertical axis is linear in depth rather than linear in time (with the depth corresponding to the peak of a parabola for a scatterer at that depth). This allows a human operator to easily position and potentially size these defects, if that is preferable to further automated analysis.

waves has also been attempted and was found to provide equivalent performance; the output is not shown here as the results are very similar and add nothing further to the demonstration of the technique.

Multiple $(c_{emission}, c_{reception})$ pairs can be filtered for, but if one wave mode is much stronger (such as the shear wave mode in the experimental data shown here), best performance has been observed if only a single pair is used. Due to the similarity of parabolas at deeper depths, strong indications can, in some cases, be observed for the incorrect wave mode. For example, a shear wave parabola will give a strong signal if (c_{shear}, c_{shear}) is used, but will also give a (slightly weaker) signal if $(c_{compression}, c_{compression})$ is used, representing a defect at a deeper depth than is actually the case. Using $(c_{compression}, c_{shear})$ would give a signal at the wrong depth and offset to the reception side and $(c_{shear}, c_{compression})$ would give a signal at the same wrong depth but offset to the other side (such as how the example parabolas in figure 7.2b look similar to an offset version of those in figure 7.2a); again, both of these outputs would be weaker than the correct output due to the inferior correlation of the filter with the actual parabola in the B-scan, but which was correct would be ambiguous if one response was not clearly larger.

7.4 Combining B-scans by shifting and scaling

The operator does not want to view multiple B-scans as output unless absolutely necessary, and hence it is desirable to combine multiple B-scans into a single output whenever possible. The effectiveness of this improves if more transducers are used, covering different angles and depths to provide good coverage of the entire area of interest. Consider the case for which the emitting and receiving transducers are in a B-scan configuration, with the transducers in a line parallel to the scan direction. In a single row of transducers, there is a single emitter, and multiple receivers, and the scan increment is constant for all transducers. The emitter does not need to have the same separation from each receiver, as long as the separation is constant over the scan. Multiple rows of transducers acting independently can also be combined (figure 7.9). Both the shift and scale operations are relatively simple.

7.4.1 B-scan scaling

For each emitter-receiver pair, the times are converted to depths by assuming a scatterer could be at any point along the emitter-receiver centre line. Any other line that extends purely along the depth axis could be used, and then this would become the line which was used in the shifting stage later, but for simplicity here, the centre line is considered. For every point along this line, the time taken, t, for a wave to travel from the emitter at speed c_1, to a point on the line, and then to the receiver at speed c_2, is calculated. Half the emitter-receiver separation is l, and the depth being considered is d:

$$t = \left(\frac{1}{c_1} + \frac{1}{c_2}\right)\sqrt{d^2 + l^2} \tag{7.7}$$

$$d = \sqrt{t^2\left(\frac{1}{c_1} + \frac{1}{c_2}\right)^{-2} - l^2} \tag{7.8}$$

If the wave does not mode-convert at the scatter point, then this reduces to:

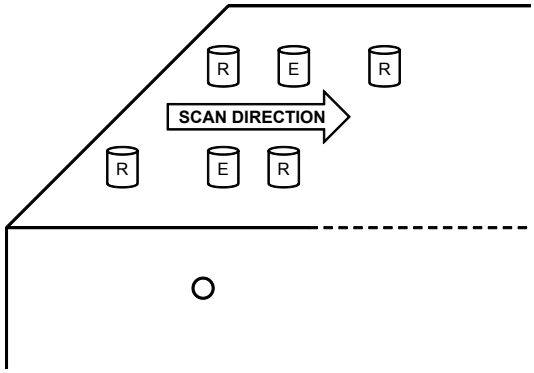


Figure 7.9: There are two rows of transducers, offset in a direction perpendicular to the scan direction, each row containing a single emitter and two receivers. The two rows are, in this case, operating independently, but follow the same scan direction. Two B-scans are formed per row (one per receiver). By shifting the B-scans such that emitter-receiver centre points overlap, and scaling such that for a single wave mode or mode-conversion, the B-scan time axis can be converted to depth for scatterers along the emitter-receiver centre line (extending into the depth of the sample), all the B-scans can be combined. Between rows, defects that extend through the sample, such as a side-drilled hole, will appear in B-scans from both rows and will be enhanced in the combination. More localised defects will be enhanced when combining a row, but then suppressed to some extent when combining multiple rows.

$$d = \sqrt{c^2t^2/4 - l^2} \tag{7.9}$$

In either case, this allows times to be converted to a linear range of depths by linear interpolation.

7.4.2 B-scan shifting

A reference grid is constructed into which data can be entered. The grid y-axis is the full linear depth range, and the x-axis is the full range of scan positions that were scanned by all the B-scans, with a scan increment equal to the scan increment used for the individual B-scans, which, as stated earlier, must be identical for all the B-scans. There is also a z-axis, with each point on the z-axis being for a different B-scan; this provides more options regarding how the data will finally be combined later. For each B-scan, the scaled and shifted data is then entered into the reference grid (at the appropriate point on the z-axis) by linear interpolation. If for a given pair, there is no information at a particular depth as it lies outside the time range for which data is available, zeros are placed in the region instead (or the data is simply not entered). Zeros are also placed (or the data not entered) for any scan position not covered by a given B-scan. Finally, the data is averaged, but only scan sections that contained data are included; regions outside the time range for which data was collected for a given B-scan and regions that were not scanned during a given B-scan are excluded. The averaging can be via taking the mean or the median, since each B-scan had its own position on the z-axis, ignoring elements for which no data was entered as already stated. Any sections of the reference grid that had no data entered by any B-scan are left as zeros. If the option to take the median is not required, and the mean is sufficient, then it requires less memory to reduce the z-axis to a single point and perform a simple addition for all data entered. If the number of entries into each individual grid element are counted, the mean can be taken simply by dividing by this second grid, conveniently dealing with regions that have not been contributed

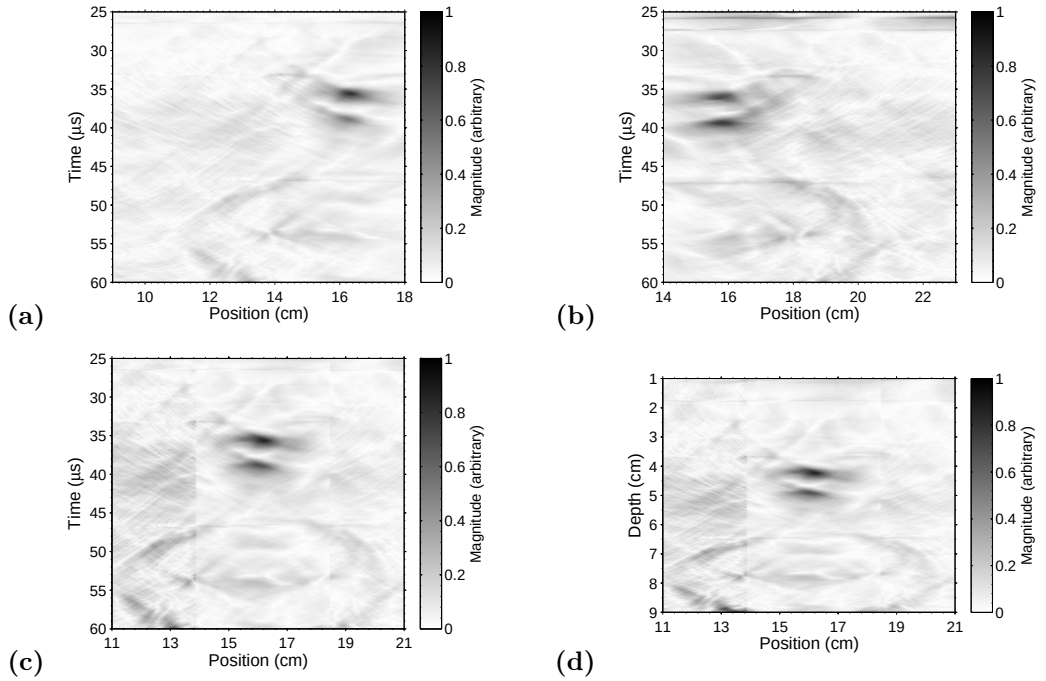


Figure 7.10: Two individual B-scans from receiving EMATs either side of an emitting EMAT, (a) and (b), are shifted to produce a single B-scan (c). Only shear waves are considered in this case, so no scaling was required. The time can still be simply converted to depth (d). The two clear indications are from the same defect (a side-drilled hole), the first reflecting from the top, and the second creeping round the bottom. This technique offers more advantage if there are multiple wave modes, as all the results can be easily combined.

to by every B-scan. An example of two individual B-scans being combined into a single B-scan is given in figure 7.10.

7.4.3 Combining B-scans from multiple rows of transducers

If scan motions between rows of transducers are parallel but offset perpendicularly, the B-scans are combined as if they were parallel with no offset. The result is that defects that are not present for both offsets will be reduced in magnitude. A defect like a side-drilled hole that is present in both scans will be enhanced as usual, as will any long cracks (since tip diffraction will occur along the length of the crack). Smaller volumetric defects will be less well-served by this technique. To tackle this situation to some extent, also output is a overlay of the B-scan showing for which perpendicular offset (corresponding to a row of transducers) each point in the combined B-scan has its maximum magnitude. The user can then see, at a glance, the approximate offset of any defects present. This only works effectively if in any given region of the B-scan there is a single clear defect.

7.5 Future work

The algorithm should be extended to better incorporate scatterers that are not in the scan plane. In addition, scatterer peak detection could be automated and then pairs of peaks used to estimate defect size and orientation. Global thresholds are commonly used to detect peaks in a HT [178],

and a similar technique could be used here. The task of splitting the peaks into pairs is non-trivial if the scatterers are not clearly spatially separated, and the separation may be ambiguous. If the separation can be achieved (which it certainly can for simple cases), measuring the size and orientation of defects is then just simple geometry which can be easily computed.

7.6 Conclusions

A B-scan contains parabolic features characteristic of a point-like scatterer, which can be effectively identified using human vision. The parabola matched filter technique, developed to perform the task of transforming parabolas to peaks, is simple to implement. It relies upon the parabolas having a temporal width such that the shape of the matched filter can change to some extent without moving outside the parabola to be filtered, and a spatial width sufficient for the matched filter to be able to average a significant number of points. The allowance for some change in shape permits a small number of matched filters to effectively filter a large range of depths, and hence the procedure can be relatively fast (of course this is dependent on the implementation). The peaks, that are the response to the matched filter having found a parabola, have a finite size, dominated by the temporal width of the feature. The matched filter also reduces the magnitude of random noise, and to some extent non-parabolic features, that are present in the B-scan. After filtering, the scan can be displayed as depth versus lateral position.

Operation of the technique against experimental and simulated data of parabolas formed by shear waves scattering from defects has been demonstrated. The technique worked very well in the case of the simulated data, as might be expected; the parabolas were spatially long, and this meant that the filter has a significant length of parabola over which to operate such that AWGN is reduced significantly. This suggests that detection of spatially long parabolas could be performed using this method, even in the presence of considerable noise. In the experimental case, although the parabolas had the same temporal width as the simulated case, they were spatially very short (not much of the parabola was present in the scan), and this reduced the effectiveness of the parabola matched filter. Despite this, the SNR showed a good improvement, the parabolas were still formed into peaks, and the depth could easily be read from the scale, which is a useful function for users not adept at assessing the depth of a scatterer based on a parabola.

Although not demonstrated here, the technique also works against parabolas formed through mode-conversion at the scattering point. Care must be taken not to confuse a parabola formed by this process with one formed without mode-conversion, as the response to the indication is shifted towards the side in which the slower wave is present in the mode-converted case, and also appears at an incorrect depth. If the signal is temporally quite sharp or the parabola spatially quite long, this would not be a serious issue, as the difference between the standard and mode-converted parabolas would mean that the indication in the mode-converted parabola matched filter response would be significantly smaller than the response from the standard parabola matched filter. For temporally wider signals that are spatially short, the difference can make the correct response ambiguous.

Scatterers not exactly in the plane perpendicular to the surface containing the emitter and receiver can still be filtered using this technique. However, the response will be defocused and smaller than for a scatterer in the plane. Scatterers at a considerable perpendicular distance from the plane would not be filtered correctly, but this is mitigated to a large extent by the emitter and receiver used for these tests having a main beam that is primarily in the plane.

7.7 Use in development of TOFDI

This chapter has enabled the removal of a considerable amount of incoherent interference from ultrasonic B-scans, following the removal of a significant amount of coherent interference by removing direct and reflected waves, as described in chapter 6. Although the removal of incoherent noise is not required or performed for the cross-sectional imaging stage, described in chapter 8, the parabola matched filter is a useful visualisation tool for B-scans (with multiple B-scans being optionally combined). B-scans are output from TOFDI, and this provides a quick way to identify and position defects, that have been transformed from parabolas in the B-scan, to points. Although this technique is not required for the overall operation of TOFDI, it does allow the operator to retain the B-scan format whilst still getting an output resembling a cross-sectional image.

Chapter 8

Time of flight diffraction and imaging

This chapter has been published [142, 143, 163] in a modified form in conference proceedings.

This chapter describes a method for converting B-scans to cross-sectional images of the sample bulk, with multiple B-scans able to form a single cross-sectional image that has a superior SNR to the SNR available from any single B-scan. This can be considered as a more general form of the parabola matched filter, introduced in chapter 7. The cross-sectional image is easy to interpret, peaks are automatically identified and positioned, with the results logged. The cross-sectional image and the list of defects are the primary outputs of TOFDI. Results using simulated and experimental data are shown, and several issues discussed, including the use of a sparse array of transducers, the presence of multiple wave modes, the importance of converting the A-scans to analytic form, and a comparison with the TFM.

8.1 Introduction

Time of flight diffraction and imaging (TOFDI), based on time of flight diffraction (TOFD), uses a sparse transducer array (laser-EMAT for non-contact applications), and adds an imaging stage based on multiple wave mode (and mode-converted) ultrasound scattering by defects. As with TOFD, the basic principle is that an ultrasound pulse emanates from the emitter, propagates through the sample bulk, scatters off the defects, and finally arrives at a receiver. Defects appear as parabolas in the B-scan, with different wave modes and defect depths leading to different parabola curvatures. Using the time-of-flight, simple geometry, and the wave speed, allows precise and accurate crack size assessment. TOFD was designed to work with a single wave mode, originally compression waves, and hence assumes there is no interference from other wave modes. TOFD also typically uses probes angled towards the expected defect position.

A sparse transducer array, typically a single emitter and multiple receivers (rather than the single receiver of TOFD), can be scanned over a sample surface with irregular geometry, and the array layout can be changed to optimise for certain defect positions (figure 8.1). Conventional phased arrays, and those using FMC, generally use piezoelectric transducers that can send and receive efficiently. TOFDI cannot make this assumption, since the transducers suitable for use in hostile conditions may only send or receive (lasers), or may be optimised for send or receive (EMATs). In addition, a continuously moving sample could leave insufficient time for FMC.

Consider a wave that has come from an ultrasound emitter, been scattered by a bulk defect, and has then reached an ultrasound receiver; the scatterer could lie anywhere on what is sometimes

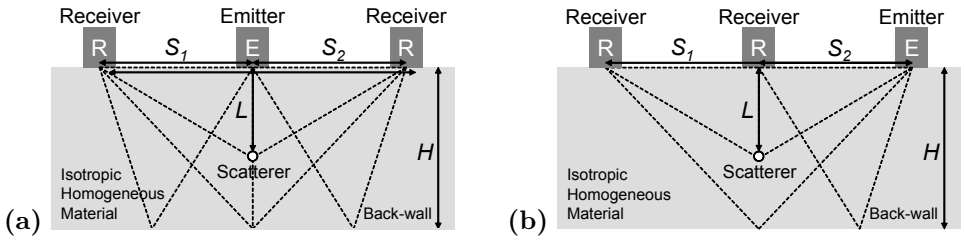


Figure 8.1: Example configurations for TOFDI. The relative position of the emitter and receivers can be varied depending on the characteristics of the defects being searched for. One or more receivers can be used, with no practical upper limit. For the work done within this project, the first configuration (a) was used rather than the second (b), but this was due to the dimensions of the sample and the necessity of the emitter and receivers to have some sample below them.

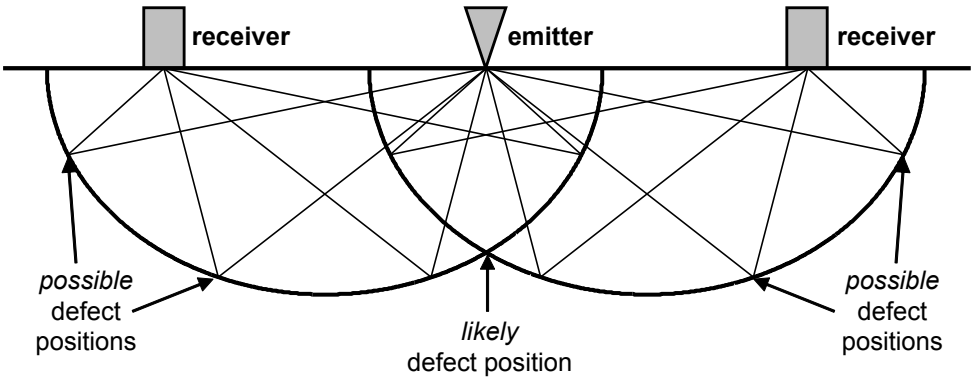


Figure 8.2: The sparse array can be used to image the sample. The standard method works on the principle that all possible defect positions lie on an ellipse defined by the time of arrival and the positions of the emitter and receiver, and that overlapping defects will highlight a likely defect position.

referred to as the travel-time locus curve. Assume a homogeneous isotropic medium and no mode-conversion. With two receivers (or two positions of the same receiver), using two ultrasound returns from the same feature, the intersection of two ellipses represents the position of that feature [1, 14], as depicted in figure 8.2. A hyperbola can also be drawn by considering the time difference between the receivers for the same feature. Three receivers (or three positions of the same receiver) allow either for redundant ellipses, or for three hyperbolas by considering the time difference between the receivers for a given feature. TOFDI, like SAFT, locates scatterers by summing all contributions for a point scatterer; if the space to be imaged is formed into a grid, each point along the travel-time locus has the signal at that time added to it. Positions that are not the correct scatterer point receive a much smaller magnitude in this process [42]. Other work has compared the use of ellipses, circles (with an extra assumption regarding the defect orientation), or optimisation techniques where cost functions based upon the arrival time data are used [184]; some such optimisation functions can be highly sophisticated [185].

As in other work of this sort [42], the emitter is assumed to approximate a point source, and the receivers are also assumed to approximate points. Scatterers are assumed to be point sources (although once the image has been produced, more sophisticated analysis could be applied), which implies that the underlying defect is voluminous or crack-like.

8.2 Theory and method

8.2.1 Pre-processing

Assuming a B-scan input (an A-scan can be used as the smallest possible B-scan), the basic processing stages have already been described in the preceding chapters. Each A-scan is time-gated to remove anything definitely not of interest, and then time-shifted so that the pulse emission can be considered to be at its central point when $t = 0$. Either a matched filter or a band-pass filter is applied to each A-scan to remove noise whilst retaining the signal of interest. Horizontal and angled linear features are removed from the B-scan, as these represent surface waves that contain no useful information regarding bulk defects, but unfortunately have a greater magnitude than the ultrasound scattered by the bulk defects. The band-pass filter applied earlier is applied again to each A-scan, to remove any out-of-band noise introduced by the removal of the linear and angled features from the B-scan. The reliability of each part of the B-scan is assessed, and a suppression is applied relative to that reliability, but this is only used for display of the B-scans, unless explicitly stated that it is also used for the cross-sectional imaging, as it can remove useful information if applied blindly. Parabola matched filtering is not applied, as this is an alternative to the full cross-sectional imaging that is to be applied here. These stages can be switched on or off as required. The analytic signal is obtained for each A-scan at the end of this pre-processing, and then the full (now complex) B-scan is passed into the cross-sectional imaging algorithm.

8.2.2 Ellipsoids

Although the equations for ellipsoids are not used to implement the cross-sectional imaging, it is the basis on which the imaging works, and hence they are described here. An ellipsoid is a locus of points, $\mathbf{P}$, in a plane, with the sum of the distances to two fixed foci, $(\mathbf{F_1}, \mathbf{F_2})$, a constant, l:

$$l = |\mathbf{F_1} - \mathbf{P}| + |\mathbf{F_2} - \mathbf{P}| \tag{8.1}$$

The total emitter-scatterer-distance, l, is a function of the position of the scatterer, $\mathbf{P}$, the position of the emitter, $\mathbf{F_1}$, and the position of the receiver $\mathbf{F_2}$. By converting l to a time by dividing by the propagation speed, the ellipsoid calculated for a given time of flight will have the defect at one of its loci. The ellipsoid centre is at (x_0, y_0, z_0), and its maximum extent in each direction is (a, b, c). The ellipsoid loci can be calculated using:

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} + \frac{(z - z_0)^2}{c^2} = 1 \tag{8.2}$$

In parametric form for $0 \leq \theta < 2\pi$ and $0 \leq \phi \leq \pi$:

$$x = x_0 + a\cos\theta\sin\phi \tag{8.3}$$

$$y = y_0 + b\sin\theta\sin\phi \tag{8.4}$$

$$z = z_0 + c\cos\phi \tag{8.5}$$

The distance from the ellipse centre to either focus is $d = \sqrt{a^2 - b^2}$. Make $b = c = \sqrt{a^2 - d^2}$

as the ellipsoid should be rotationally symmetric about the x-axis in this case (there is no reason to favour the y-axis over the z-axis).

Assume the foci are at **A** and **B** (the emitter and receiver for this work) and the centre of the ellipsoid is at $\mathbf{C} = (x_0, y_0, z_0)$:

$$d = |\mathbf{B} - \mathbf{A}| /2 \tag{8.6}$$

$$\mathbf{C} = (\mathbf{B} + \mathbf{A})/2 \tag{8.7}$$

The ellipsoid equations so far have implicitly assumed that the foci both sit along the x-axis. If this is not true, the axes can be rotated to fit the assumption using a simple rotation matrix [186,187]. If the ellipsoid has an initial direction, $\hat{\mathbf{x}}$, this must be rotated to $\hat{\mathbf{e}} = (\mathbf{B}-\mathbf{A})/|\mathbf{B}-\mathbf{A}|$. The order in which the two rotation matrices (pitch and yaw, roll is not required due to the ellipsoid being assumed rotationally symmetric) are applied is not important, as the ellipsoid is rotationally symmetric about the x-axis; however, although the vector (the local $\hat{\mathbf{x}}$) will be pointing in the desired direction in either case, the direction of the other axes (the local $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$) will differ between the two rotations. The rotation matrices are given below, with the axis about which the rotation occurs given as the subscript:

$$\mathbf{R_x}(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \tag{8.8}$$

$$\mathbf{R_y}(\beta) = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \tag{8.9}$$

$$\mathbf{R_z}(\gamma) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8.10}$$

Of course, matrix multiplication is defined as:

$$(\mathbf{AB})_{i,j} = \sum_{r=1}^{n} \mathbf{A}_{i,r}\mathbf{B}_{r,j} \tag{8.11}$$

The rotation matrices are then (depending on the order):

$$\mathbf{R_{zy}}(\beta,\gamma) = \begin{bmatrix} \cos\beta\cos\gamma & \sin\gamma & -\sin\beta\cos\gamma \\ -\cos\beta\sin\gamma & \cos\gamma & \sin\beta\sin\gamma \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \tag{8.12}$$

$$\mathbf{R_{yz}}(\beta,\gamma) = \begin{bmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ -\sin\gamma & \cos\gamma & 0 \\ \sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix} \tag{8.13}$$

These are related by:

$$\mathbf{R_{zy}}(\beta,\gamma) = \mathbf{R_{yz}^T}(-\beta,-\gamma) \tag{8.14}$$

Rotations must be done before offsets are applied (x_0,y_0,z_0), as otherwise the offsets are rotated as well, which is not the desired behaviour. If the matrix describing the position of points of the ellipse is represented by $\mathbf{Q}$, and the desired output after rotation and translation is $\mathbf{P}$, then (arbitrarily choosing $\mathbf{R_{yz}}$ as the rotation matrix):

$$\mathbf{Q} = \begin{bmatrix} a\cos\theta\sin\phi \\ b\sin\theta\sin\phi \\ c\cos\phi \end{bmatrix} \tag{8.15}$$

$$\mathbf{P} = \mathbf{R_{yz}Q} + \mathbf{C} = \begin{bmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ -\sin\gamma & \cos\gamma & 0 \\ \sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix} \begin{bmatrix} a\cos\theta\sin\phi \\ b\sin\theta\sin\phi \\ c\cos\phi \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \tag{8.16}$$

If the problem can be reduced to two dimensions, it becomes the computationally faster:

$$\mathbf{P} = \mathbf{R_z Q} + \mathbf{C} = \begin{bmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} a\cos\theta\sin\phi \\ b\sin\theta\sin\phi \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{8.17}$$

The required rotation angle is taken from the rotation necessary to go from $\hat{\mathbf{x}}$ to $\hat{\mathbf{e}}$.

$$\hat{\mathbf{e}} = \mathbf{Q}\hat{\mathbf{x}} = \begin{bmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ -\sin\gamma & \cos\gamma & 0 \\ \sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{8.18}$$

$$\hat{\mathbf{e}} = \begin{bmatrix} \cos\beta\cos\gamma \\ -\cos\beta\sin\gamma \\ \sin\beta \end{bmatrix} \tag{8.19}$$

Since $\hat{\mathbf{e}}$ is a known vector, it is trivial to solve for γ and β.

8.2.3 Mode converted ellipsoid equivalents

If the wave is mode-converted at the intermediate point, compression to shear or shear to compression (assuming a bulk wave in an isotropic homogeneous sample), the wave speed is not constant during the time of flight, and the intermediate points can no longer be described by an ellipse. An analytic description of the position of the intermediate points is still possible.

Consider a total time of flight, t, split into two sections, before the intermediate point, t_1, and after the intermediate point, t_2, such that $t = t_1 + t_2$. The wave from the emitter travels $r_1 = c_1 t_1$ before reaching the intermediate point, where c_1 is the wave speed of the first wave mode; all the positions it could travel to are defined by a circle of radius r_1 centred at the emitter. At the intermediate point, mode conversion occurs to the second wave mode, moving at a speed of c_2. After exactly t_2 (the remaining time, $t_2 = t - t_1$), the receiver will receive any wave at a distance equal to $r_2 = c_2 t_2$; all the positions the wave could travel from are defined by a circle of radius r_2 centred at the receiver. The emitter and receiver can be considered as foci (analogous to the elliptical case), and the intersection points of the two circles centred at the foci define the possible intermediate points.

With reference to figure 8.3, the relevant equations are:

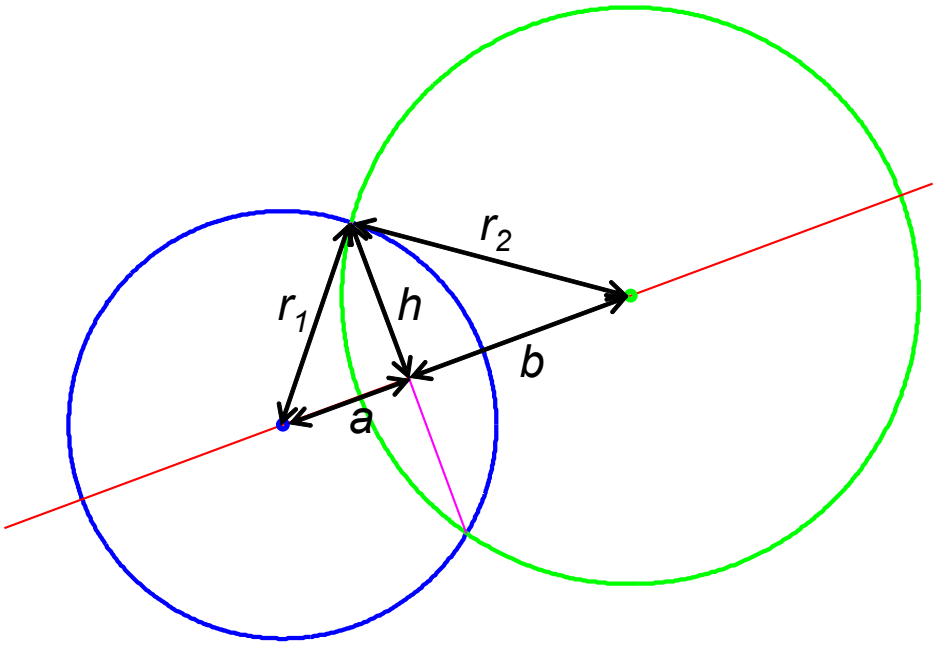


Figure 8.3: The intersection of two circles. The distance from the foci to the intersection point is r_1 and r_2 from each circle. These distances can be split into parallel and perpendicular components (relative to the line of symmetry through both circles). The perpendicular distance, h, is the same for both r_1 and r_2, whereas the parallel distance varies, a and b, respectively.

$$r_1^2 = h^2 + a^2 \tag{8.20}$$

$$r_2^2 = h^2 + b^2 \tag{8.21}$$

The total separation of the foci, d, is, for foci at vector positions $\mathbf{p_1}$ and $\mathbf{p_2}$ respectively:

$$d = a + b = |\mathbf{p_2} - \mathbf{p_1}| \tag{8.22}$$

Substitution reaches a solution for a and h:

$$a = \frac{r_1^2 - r_2^2 + d^2}{2d} \tag{8.23}$$

$$h = \sqrt{(r_2^2 + r_1^2)/2 - (r_2^2 - r_1^2)^2/4d^2 - d^2/4} \tag{8.24}$$

Then the intersection points are given by:

$$\mathbf{p} = \mathbf{p_1} + \left(\frac{\mathbf{p_2} - \mathbf{p_1}}{d}\right) a \pm \left[\left(\frac{\mathbf{p_2} - \mathbf{p_1}}{d}\right) \times \hat{\mathbf{z}}\right] h \tag{8.25}$$

The $\pm$ allows for both intersection points to be calculated, remembering the system is symmetrical about the line between the foci (this applies to rotational symmetry about the line in three dimensions).

The equations work in three dimensions as well as two, except for the cross product with $\hat{\mathbf{z}}$.

That is used as a unit vector perpendicular to the unit vector $\left(\frac{\mathbf{p_2}-\mathbf{p_1}}{d}\right)$, assuming that vector is in the x-y plane. If the vector is not in the x-y plane, the world axes can be rotated until it does lie in the x-y plane (rotation matrices can be used), and then rotated back again after finding the intersection points.

There are infinite solutions if:

$$d = 0,\ r_1 = r_2 \tag{8.26}$$

This is because the circles are coincident; the equations for a and h break down in this case. However, there are no solutions if:

$$d > (r_1 + r_2) \tag{8.27}$$

$$d < |r_2 - r_1| \tag{8.28}$$

The first case is because the circles are too far apart to touch. The second case is because one circle completely contains the other. Both result in no intersections (h becomes imaginary). There is one solution if the circles are just touching:

$$d = (r_1 + r_2) \tag{8.29}$$

$$d = |r_2 - r_1| \tag{8.30}$$

There are two solutions if otherwise:

$$|r_2 - r_1| < d < (r_1 + r_2) \tag{8.31}$$

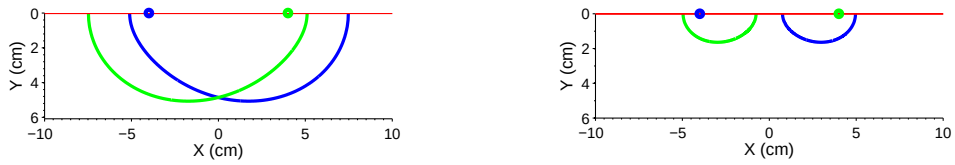
The result of this is given for a selection of t values in figure 8.4, in which the potential scatterer positions define an oval shape. The blue line is the intersections if the blue focus emits/receives the faster wave. The green line is the intersections if the green focus emits/receives the faster wave. The red line passes through both foci. When the total time is large, solutions exist when t_1 or t_2 is close to t, and at very large t, the solutions become more circular. When the total time is small, solutions only exist when the faster wave travel time is close to t. There is a single elliptical solution if both wave speeds are the same.

8.2.4 Directivity and distance compensation

The angle to each grid element can be calculated knowing the position of the transducer, (x_0, y_0, z_0), and the grid point, (x, y, z). Using spherical coordinates (but switching y and z to make it easier to change between two and three dimensions):

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \tag{8.32}$$

$$\theta = \cos([y - y_0]/r) \tag{8.33}$$



(a) **(b)**

Figure 8.4: Potential defect positions for a mode-converted wave, formed through circle intersections. The blue line is the intersections if the blue focus emits/receives the faster wave. The green line is the intersections if the green focus emits/receives the faster wave. The red line passes through both foci and here represents the surface. The transducers are at -4cm and 4cm, the wave speeds are 5960m/s and 3235m/s, and the total time of flight is 30μs (a) and 18μs (b).

$$\phi = \tan([z - z_0]/r) \tag{8.34}$$

In this case, the depth is y, and if the system is two dimensional, then $z = z_0$ and $\phi = 0$. Using calibration data, it is possible to create a map of how the directivity affects the signal magnitude for each transducer. Distance, r, can also be compensated for at this stage, as it has already been calculated to obtain the angle:

$$r = \sqrt{(x - x_e)^2 + (y - y_e)^2 + (z - z_e)^2} + \sqrt{(x - x_{r_n})^2 + (y - y_{r_n})^2 + (z - z_{r_n})^2} \tag{8.35}$$

As discussed in section 2.1.5, the appropriate correction for a point in three dimensional space is $A = A_0 r_0/r$ and for a point in two dimensional space is $A = A_0\sqrt{r_0/r}$. The reference distance, r_0, is the distance at which the amplitude is equal to its reference value, A_0, as measured by a reference receiver.

This is acceptable when comparing receivers with equally sized apertures (and other identical properties) at r_1 and r_2. Some workers [38, 66] consider the signal level to be $A = A_0/\sqrt{r_e r_r}$ after scattering and reception (a simple multiplication and assuming $r_0 = 1$), but it can become more complicated after scattering. Consider a perfect reflector parallel to the surface plane on which the emitter and receiver are placed, that specularly reflects all energy directed towards it. If the emitter has a directivity such that most of the incident energy impinges on the perfect reflector, then $A = A_0\sqrt{r_0/(r_e + r_r)}$ in two dimensions, ignoring edge effects such as diffraction and mode conversion. If the reflector is half the size, then $A = \frac{1}{2}A_0\sqrt{r_0/(r_e + r_r)}$. The reason for considering this reflector is to demonstrate that $A = A_0 r_0/\sqrt{r_e r_r}$ is not always valid. Amplitude after scattering is dependent on the nature of the scatterer and its size, and issues such as angled defects, waves that creep round the defect, rough defects, concave defects, etc. have not been touched upon. Intrinsic attenuation does not have this issue, and always combines as $A = A_0 e^{-\alpha(r_e + r_r)}$.

As an example, a parabola in a B-scan collected during the testing of TOFDI, can be observed clearly over approximately 50mm (a deviation from the centre point of 25mm), for a defect depth of approximately 45mm. For an emitter-receiver separation of 80mm, the change in signal level purely from changes in distance (assuming $A = A_0 r_0/\sqrt{r_e r_r}$), is merely 3.5% over the length of the

parabola. If it were observed over 80mm (such that the emitter or receiver are directly over the defect at the extremities), the change would still only be 14%. Since the error due to the defect being something other than an omni-directional scatterer is likely to be at least of this size, making any correction for distance in this way is of dubious value in improving the imaging. There are additional problems if there are multiple defects or mode conversions, such as the extremity of one parabola overlapping with the centre of another, and the centre of that parabola being erroneously enhanced and combined with the other parabola, and this makes the distance correction all the more perilous. If a large depth range is being scanned, and it is desirable to adjust for distance down the depth, then this can be done to the final imaged output; although not as accurate, it does avoid a lot of problems. Multiple defects and mode conversions causing overlapping parabolas also causes an issue with the directivity compensation.

There are other problems with compensating for angle/distance on a per-scan basis, rather than on the final image. There are angles/distances at which the signal is expected to be small. If this is compensated for by increasing the gain for signals expected to be from those angles/distances, assuming a constant level of noise, the algorithm effectively gives poorer quality parts of the signal (lower SNR) equal importance to good quality parts of the signal (high SNR). This makes the image output worse in poor SNR situations. Indeed, it would seem sensible to give less weight to low SNR parts of the signal, which naturally occurs unless prevented by the compensation.

Finally, intrinsic attenuation could be compensated for, but in addition to the criticisms above applying, it has the additional problem that it can be quite difficult to get an accurate measure of the attenuation factor for an ideal set-up, and in actual usage, the factor would have to be calculated on-the-fly for a sample that may have a rough surface and whose front and back sides cannot be guaranteed to be perfectly parallel (not to mention the need to incorporate directivity, geometric attenuation, reflection losses including mode conversion, etc.). This makes an accurate estimation of the attenuation factor very difficult.

As a result of these issues, compensation for distance (geometric and intrinsic), and directivity, is implemented within the cross-sectional imaging algorithm, but it is not typically used.

8.2.5 Signals from the top and bottom of a side-drilled hole

When dealing with a crack-like defect, there is clear diffraction from the top and bottom of the crack, and these diffracted signals appear in the B-scan. Much of the testing in this project used a side-drilled hole, and with the sort of geometry used, the bottom of the defect is not directly illuminated. Despite this, the B-scan and the output image appear to show a signal from the top and bottom of the hole. To explain this, a very simple two-dimensional model was constructed that modelled a perfect side-drilled hole, and point-like transducers. The model finds the range of lines that intersect a circle (a side-drilled hole) starting at a point-like emitter, and this defines the parts of the hole that are illuminated. The same calculation is performed for the range of lines from the point-like receiver to the hole, and this defines the parts of the hole that the receiver can detect signals from in this simple model. The incident and outgoing lines can be seen in figure 8.5a. Waves are then allowed to creep round the defect (at the same bulk wave speed), continually re-emitting waves by the Huygens-Fresnel principal, and if one of these points is within the range that can be detected by the receiver, then that part of the wave is detected. For each incident point and each re-emission point that the receiver can detect, the total time of flight is calculated, and the results of this can be seen in figure 8.5b. Note that there is a considerable contribution in

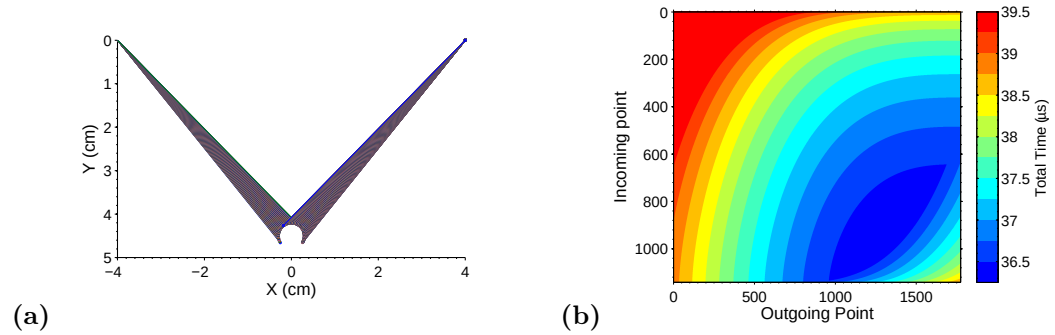


Figure 8.5: The range of directly illuminated incident points and detectable outgoing points is calculated (a), and the waves are allowed to creep round the scattering circle before re-emission at one of the outgoing points that can be detected by the receiver. The total time of flight is calculated (b), and in this case has been split into thirteen groups with the first starting at 36.25µs and the last ending at 39.5µs, each region being 0.25µs wide, so that regions of similar time can be clearly seen. The wave speed is assumed constant at 3200m/s. Equivalent times for a point scatterer at the top and bottom of the circle are 36.25µs and 39.05µs respectively.

the figure from the bottom of the time scale, and another one from the top of the time scale. This goes some way to qualitatively explain the "top" and "bottom" of the side-drilled hole observed in the results.

The signals from horizontal defect segments are strong as there is little variation in path length over the horizontal segment (for a given emitter-defect-receiver geometry); the signal contributions vary little in phase and hence constructively interfere. The signals from vertical defect segments are weak due to the considerable variation in path length over the vertical segment; the signal contributions vary significantly in phase and hence destructively interfere. In general, a particular defect edge will give a strong signal if it follows a path of an isochrone (a trajectory of equivalent time, an ellipse) [1]. Defects which have a similar shape to the ellipse will not be isochronic except when in what is known as the "glint" position, when the defect actually is approximately parallel to the isochrone. At other offsets, there will be sufficient destructive interference to make the signal very weak [1].

Waves must creep around curved surfaces by a process of continual diffraction (which also radiates back into the bulk), and although these creeping waves are not surface waves (no exponential decay of amplitude), the speed of the waves can differ from that of the bulk waves [2]. For larger radius curvatures, the speed difference is small, and for smaller radius curvatures, the distance travelled at this speed is minimal, and hence the error introduced is relatively small in many cases. Although the error may be minimal, as it has not been properly assessed, this model is completely insufficient to be used quantitatively. Additionally, no consideration is made to the amplitude of the arriving signals, only the time of arrival. In the more general case, FEA is used.

8.2.6 Cross-sectional imaging

It is assumed that the scan direction is constant, and that the emitter and receivers are static relative to each other. Each emitter-receiver pair is assigned its own imaging boundaries. This is the area over which, for a given emitter-receiver position, the pair could realistically detect a signal scattered from a defect. The simplest method to determine the boundary is through experimentation. These boundaries are static with respect to the emitter-receiver centre position (and

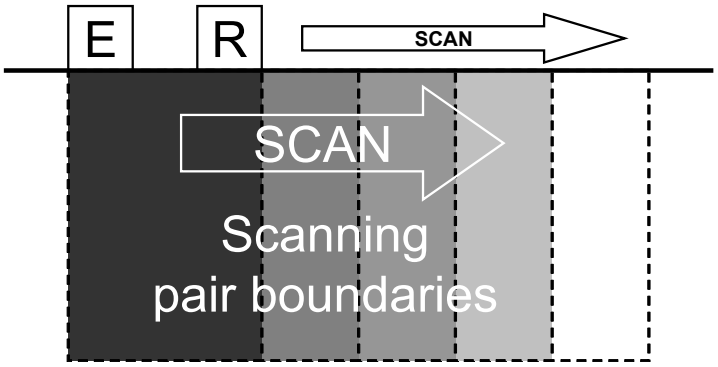


Figure 8.6: The emitter-receiver pair has an imaging boundary, the area (determined empirically) it can image at any given A-scan position. As the pair move with the scan, the imaging boundary moves with them. The total area covered by the moving pair boundaries defines the total imaged area.

the orientation remains in the same direction relative to the constant emitter-receiver separation vector). As the scan moves, the boundaries move with the emitter-receiver pair, as shown in figure 8.6. The size and position of the boundaries are chosen based on calibration data and knowledge of the region over which parabolas are likely to be observable in the B-scan. Each emitter-receiver pair can have a different pair boundary, in both position and extent.

Considering the extremes of these boundaries, over the entire scan, and for all possible emitter-receiver pairs, defines the maximum imageable area which forms the imaging grid. Outside of this area, it is unrealistic to expect any ultrasound scattered off defects to be useful in detecting defects.

The imaging grid is split into elements defined by the grid resolution, and the same grid resolution is used to split the pair boundaries into elements. The grid resolution is chosen such that the spacing, g, is no greater than a tenth of the shortest wavelength, λ_{min}. This requires the slowest wave speed, c_{min}, and the highest frequency, f_{max}:

$$g \leq 0.1 c_{min}/f_{max} \tag{8.36}$$

In this case, the grid spacing can be no larger than 0.2mm, for a wave speed of 3226m/s, and for numerical convenience, a frequency of 1.613MHz (in reality the maximum frequency is nearer 1MHz). Such a condition on the grid spacing means that small changes in the signal can potentially appear in the cross-sectional image.

Each emitter-receiver pair has a scalar separation between the transducers, and a pair boundary. Considering a static position (a single A-scan), for each unique separation and boundary, the distance from the emitter, d_e, and receiver, d_r, to each grid element, is calculated, and the two distances are kept separate in order to deal with mode-conversions later:

$$d_r = \sqrt{(x - x_r)^2 + (y - y_r)^2 + (z - z_r)^2} \tag{8.37}$$

$$d_e = \sqrt{(x - x_e)^2 + (y - y_e)^2 + (z - z_e)^2} \tag{8.38}$$

$$d = d_e + d_r \tag{8.39}$$

The direction of the scan is also recorded, in terms of which coordinates increase as the scan

progresses, and which decrease. If two pairs have the same boundary and separation, the distances calculated can be reused, although the computational time saved by doing this is minimal. At this stage, compensation can be made for variation in magnitude expected due to distance and angle, by calculating scale factors that can be applied at the next stage, and storing them in a grid of equal dimensions to the pair boundary grid. However, compensation for distance and directivity is not typically used, as already discussed.

The distances calculated are converted to times by dividing by each wave propagation speed of interest, c, with a set of times each. There are various permutations by which these speeds can be combined; the two without mode conversion being compression or shear throughout the emitter-element-receiver flight ($c_e, c_r = c_c$ or $c_e, c_r = c_s$), and the two with mode conversion being compression or shear for the emitter-element flight, and the opposite for the element-receiver flight ($c_e = c_c$, $c_r = c_s$, or $c_e = c_s$, $c_r = c_c$). The times are combined in each speed permutation of interest, with each permutation saved in memory:

$$t_r = d_r/c_r \tag{8.40}$$

$$t_e = d_e/c_e \tag{8.41}$$

$$t = t_e + t_r \tag{8.42}$$

It is assumed that each A-scan collected data at the same time points. It is then trivial to convert the times calculated for each permutation, t, to the corresponding nearest signal time point in the A-scan, n, relative to a zero-time point, t_0, and a sampling frequency, f_s:

$$n = \lfloor (t - t_0) f_s + 1.5 \rfloor \tag{8.43}$$

The interpolation method is nearest neighbour, as can be seen from the use of the floor function ($\lfloor\rfloor$) with 0.5 added (an additional 1 is added as MATLAB's arrays start at one rather than zero). This is acceptable due to the high sampling frequency used of 100MHz or 1GHz. Even at the highest ultrasound signal frequency of interest, f, the wave amplitude received will change very little between two sampling points. The biggest change is at the zero crossing of the highest frequency sine wave (this point has the greatest differential), so considering a middle point exactly on the zero crossing, and taking a point either side, each a quarter sampling period, T, from the zero crossing, the largest error is:

$$\Delta A = \sin(2\pi f T/4) - \sin(-2\pi f T/4) = 2\sin(f T \pi/2) \tag{8.44}$$

A quarter sampling period is used either side, as the maximum deviation from the correct sampling point is half a sampling period. For the example of a 2MHz signal frequency (there is very little frequency content above this), with a sampling frequency of 100MHz, the change is approximately 3% of the peak-to-peak amplitude, and this is the maximum possible for this work. At the more dominant frequency of 0.5MHz, the maximum change is approximately 0.8%. The loss of resolution due to using this interpolation method is far less than the loss of resolution from other factors, and hence need not be considered further. Any grid points that sit outside the collected time range are set to point to a zero value.

For each A-scan position, the pair boundary is shifted to the corresponding point in the overall imaging grid based on the joint emitter-receiver position (figure 8.6). Those points in the overall imaging grid then have a corresponding time in the A-scan data. The corresponding signal points from the A-scan, for those times, are added to the existing values in the overall grid. This is an addition of complex numbers, as the A-scans were earlier converted to analytic form. As with some SAFT work [42], the real and imaginary parts of the analytic signal are used when forming the image, as this has been found to improve resolution. Using just the real part of the A-scan introduces the signal character to the cross-sectional image, so that for a bipolar signal for example, imaged points appear bipolar. Using just the magnitude widens the response, and means that two dissimilar signals appearing at the same cross-sectional image point can add, rather than subtract as they might if added including phase. Using just the phase means that low SNR parts of the image are over-emphasised. Combining magnitude and phase keeps the defect response sharp, and maintains good SNR.

Separate overall grids are kept for different emitter-receiver pairs and speed permutations, and the absolute magnitude is taken of each. As each speed permutation and emitter-receiver pair may produce a defect indication with a different phase, combining the complex grids would reduce the amplitude of the defect indication. It is better to use the analytic signal when combining different scan positions, and the magnitude when combining speed permutations and different emitter-receiver pairs. The number of entries to each point in the overall grid is counted, ignoring any points for which data could not be added as they were outside the time range for the A-scan.

The image grids for all the speed permutations are combined by summation. At this stage, compensation can be made for variation in magnitude expected due to the wave mode or conversion. Regarding combining the different receiver grids, both summing and taking the product (including the product of the number of entries for consistency when averaging) has been tried. The product can only be used if the receivers can all see the defect, as otherwise it will suppress the appearance of the defect in the cross-sectional image. In cases where it is known that not all the receivers will pick up scattered waves from the defect, those that can detect it can be multiplied together, and the others can be kept separate at this stage.

The cross-sectional image is divided by the number of entries to each overall grid element, to obtain a mean magnitude. Any elements with a number of entries below a threshold are set to zero, as those grid points are not considered reliable. If at the earlier stage, some of the images from receivers were multiplied together and some summed together, both sets of images can now be combined by summing the normalised cross-sectional images. The cross-sectional imaging grid can then be displayed. If the imaging was two-dimensional, the grid can be shown without further processing. However, if the imaging was three-dimensional, then for display purposes, the grid is usually summed along the third dimension, the dimension which is most perpendicular to the scan direction. This compresses all the images to a focus plane.

8.2.7 Defect highlighting

Several operations are performed to highlight defects in the cross-sectional image. First, a simple mean filter is applied [168], with a length of $cw/10$ (this was chosen by experimentation), where w is the ultrasound pulse width in time, and multiplying by the wave propagation speed, c, converts the time to a pulse width distance. This is then converted to a number of grid elements by dividing by the grid spacing. The size is forced to an odd number (by adding an element if necessary) so

that the filter has a centre point, and it is made into a square in two dimensions, or into a cube in three dimensions (each dimension of length $cw/10$). As the most simple convolution filter, it can be applied very quickly and easily using, for example, the standard FFT technique [156]. The problems usually associated with this filter is that a single pixel with a very unrepresentative value can change the mean value significantly, and edges become blurred. Testing has shown that these issues are not particularly detrimental in this case. If they became detrimental, a median filter could be used instead, but this was tested, and for the sorts of cross-sectional images seen, no improvement in moving to a median filter was observed. To apply the filter, a grid is created that is the same size as the cross-sectional image with the addition of half the filter size (not including the centre point) in each active dimension. As it is to be applied using the standard FFT technique, half the filter is placed in the "positive" side of filter grid, and half in the "negative" side, by setting those points to the inverse of the number of points in the filter. The filter can then be applied by taking the complex conjugate of the FFT of the filter, zero-padding the cross-sectional image to the same size as the filter and performing an FFT, multiplying the two together, performing an IFFT, and finally removing the additional zero-padded regions that were added (as was done in section 7.2.3). Any regions that were previously set to zero as they contained too few data points are again set to zero; they may have stopped being zero due to the blurring caused by the filter.

The image standard deviation, σ, is obtained, ignoring any parts that were set to zero due to insufficient data entries; any points $\leq 5\sigma$ are set to zero, and any points $> 5\sigma$ are set to one. This is known as the Rose criterion, which states that to reliably identify an object, the SNR must be better than approximately 5 [188].

Peaks are extracted from identified regions, by searching for local maxima, in a crude but sufficiently effective fashion. Each point in each of the identified regions is compared to every directly adjacent point in the region (8-connected, so diagonally adjacent points are compared), and if the point is greater than those adjacent to it, it is recorded as a local peak. If desired, only the largest magnitude peak can be identified per region, but by default, all local peaks in a given region are identified. In most cases, there will only be a single local peak in any given region due to the application of the mean filter removing high frequency local variation. The identified peaks are reported as a magnitude and a position.

8.2.8 Inverse imaging

Different defect types, such as a crack, notch, or hole, may produce different scattered signals. To get the data for a single defect, the scatterer is assumed to be point-like. Based on the identified peaks, the time, t, for the relevant speed permutations, c_e and c_r, is extracted, and converted to a sample point in a given A-scan, n, relative to a zero-time point, t_0, and a sampling frequency, f_s:

$$t = \frac{1}{c_e}\sqrt{(x-x_e)^2+(y-y_e)^2+(z-z_e)^2} + \frac{1}{c_r}\sqrt{(x-x_r)^2+(y-y_r)^2+(z-z_r)^2} \tag{8.45}$$

$$n = \lfloor (t-t_0)f_s + 1.5 \rfloor \tag{8.46}$$

This is performed for each A-scan, for each emitter-receiver pair, as the pair moves. The signal for that peak is extracted from each A-scan using the calculated time for the peak. Half the pulse is assumed to appear prior to that time, and half the pulse after that time, with a known pulse

length. Each A-scan can be associated with an emitter-scatterer-receiver angle and distance. There has not been sufficient time to study calibration defects against which to make a comparison, and it is not clear the extent to which different defects could be distinguished, if at all. Consequently, this technique has only been used for very limited testing.

8.2.9 Speed measurement

Wave speeds were measured using the standard method of back-wall reflection for compression waves and pitch-catch for Rayleigh waves. The back-wall was sufficiently perpendicular to the surface for the reflected energy to travel back to the surface without significant angular deflection. In both cases, piezoelectric transducers were used for simplicity; they can easily operate in both generation and reception modes and can be coupled using a gel couplant. The compression wave speed was measured as (5931±11)m/s and the Rayleigh wave speed was measured as (2987±12)m/s, with the linear regression including errors in both distance and time [189]. The standard values for mild steel are 5960m/s for compression waves and 2996m/s for Rayleigh waves [19], making the measured values within 0.5% (2.6σ) and 0.3% (0.75σ) respectively. The shear wave speed can be calculated from the two measured speeds, and is (3226±14)m/s, which is within 0.3% (0.64σ) of the standard value of 3235m/s.

For many cases, either standard or measured values could be used as they are so close, and where simulations have been carried out, either the standard values have been used, or where appropriate, values rounded to the nearest 100m/s or even the nearest 1000m/s.

8.2.10 Errors

The limiting errors for TOFD are covered in several papers [2, 13, 25], including errors in timing due to changes in probe separation or unaccounted for variation in the surface of the sample [13], and variations in ultrasound speed. The errors for TOFDI have been calculated using a different method, but parameters considered are based upon those used in the error modelling of TOFD [25].

Unlike piezoelectric transducers, there is no error due to coupling film thickness, probe shoe effects, and similar issues. Variations in speed are particularly of interest if the sample is known to deviate from the assumption of homogeneous and isotropic to a significant extent. It is also often the case that the sample will not be perfectly level on the scanning surface, meaning that the probes could be displaced from their assumed position in the direction perpendicular to the surface. However, this could be tracked and accounted for if a significant deviation is expected. Note that pulse width is usually defined as the time for the pulse to reduce to 10% of its peak value [3].

Greater resolution can be achieved when the travel-time loci intersect perpendicularly (from two different receivers for example), such that a small change in the time of flight results in a large change in the defect indication position. However, greater sensitivity to changes in time of flight can also result in greater susceptibility to errors caused by problems such as an incorrectly known ultrasound speed.

8.2.11 Buffer regions

There are pre-processing stages which minimise horizontal and angled features due to the presence of direct and reflected surface waves, and back-wall reflections. However, if these features are to

be avoided altogether, the collection geometry can be manipulated so that they can be removed by time-gating whilst retaining the defect signals. This does limit the areas which can be scanned effectively for defects, hence the reason the removal techniques were implemented. Consequently, these buffer regions have not been necessary.

Consider an emitter at $\mathbf{E} = (E_x, E_y)$, a receiver at $\mathbf{R} = (R_x, R_y)$, and a defect at $\mathbf{D} = (D_x, D_y)$. Add boundaries at B_x (there are separate left and right boundaries but they are both referred to as B_x) and B_y. The wave speeds of interest are the slowest bulk wave (shear wave c_s), the fastest lateral wave (compression wave c_c), and the slowest surface wave (Rayleigh wave c_R). The surface is assumed flat and defined as the zero point on the y-axis.

The time, t_c, for the fastest wave (a compression wave) to go from the emitter to the defect and then to the receiver is:

$$t_c = \frac{1}{c_c}\left(\sqrt{\left(D_x - E_x\right)^2 + \left(D_y - E_y\right)^2} + \sqrt{\left(D_x - R_x\right)^2 + \left(D_y - R_y\right)^2}\right) \tag{8.47}$$

The time, t_s, for the slowest wave (a shear wave) to make the same trip is:

$$t_s = \frac{1}{c_s}\left(\sqrt{\left(D_x - E_x\right)^2 + \left(D_y - E_y\right)^2} + \sqrt{\left(D_x - R_x\right)^2 + \left(D_y - R_y\right)^2}\right) \tag{8.48}$$

Now set $E_y = 0$ and $R_y = 0$. Consider the directly received surface Rayleigh wave (the surface compression wave will always be received first and hence can be neglected). The last surface wave arrives at:

$$t_d = \left(R_x - E_x\right)/c_R \tag{8.49}$$

Time t_c needs to be larger than the time for a direct surface wave t_d:

$$t_c > t_d \tag{8.50}$$

The earliest time a bulk wave arrives after scattering off a defect is when the defect is at the mid-point between the emitter and receiver. Define this point as $\mathbf{D} = (0, D_y)$ for simplicity (it is only a simple translation of the coordinates):

$$\frac{1}{c_c}\left(\sqrt{E_x^2 + D_y^2} + \sqrt{R_x^2 + D_y^2}\right) > \left(R_x - E_x\right)/c_R \tag{8.51}$$

Since it is now the midpoint, it is also possible to define $R_x = -E_x$:

$$D_y > R_x\sqrt{\frac{c_c^2}{c_R^2} - 1} \tag{8.52}$$

Time t_s needs to be less than the time for a surface wave reflection or a back-wall reflection t_b:

$$t_s < t_b \tag{8.53}$$

The fastest back-wall reflection arrives at:

$$t_b = \frac{1}{c_c}\sqrt{\left(R_x - E_x\right)^2 + 4B_y^2} \tag{8.54}$$

Again, making the assumption that the emitter and receiver positions are known and fixed (if they are not fixed, the extremes of position can be used):

$$\frac{1}{2}\sqrt{(c_c t_s)^2 - (R_x - E_x)^2} < B_y \tag{8.55}$$

The fastest surface wave reflection arrives at:

$$t_r = \frac{1}{c_c}|2B_x - E_x - R_x| \tag{8.56}$$

If the assumption is then made that the emitter and receiver positions are known and fixed (if they are not fixed, the extremes of position can be used), then for the right boundary, B_{x_r}, and the left boundary, B_{x_l}:

$$\frac{1}{2}(E_x + R_x + c_c t_s) < B_{x_r} \tag{8.57}$$

$$\frac{1}{2}(E_x + R_x - c_c t_s) > B_{x_l} \tag{8.58}$$

This allows appropriate boundaries to be chosen. Additional buffering should be added to compensate for the emitted wave having a finite duration, as it is desirable for the bulk wave to be completely received before observing the reflected waves.

8.2.12 Total focusing method

As a comparison to TOFDI, TFM was implemented, such that it could operate against simulated data. A one-dimensional linear 32 transducer array with a spacing of 1mm was simulated. Regarding implementation, several optimisations were made. The imaging area was split into elements with a spacing an integer divisor of the transducer array spacing, and starting such that each transducer is at the centre of an image element. Consider a transducer exactly at the horizontal centre of the imaged area, a point defined as $(0, 0)$ for convenience, with the grid extending as $(-x_{max} \leq x \leq x_{max}, 0 \leq y \leq y_{max})$. The distances, d, to the each grid element, were calculated for one side of the grid ($x \geq 0$):

$$d = \sqrt{x^2 + y^2} \tag{8.59}$$

By symmetry, the other side of the grid ($x < 0$) has the same distances to that transducer, and hence the output can be flipped and combined with the calculated side to get the full set of distances.

The transducer array has a constant spacing, x_{trans}, corresponding to an integer number of grid element spacings. The separation between transducers being used for emit and receive, for an array with M elements is ($m \in \mathbb{Z}^+$):

$$x_{sep} = m x_{trans} \tag{8.60}$$

$$0 \leq m \leq M - 1 \tag{8.61}$$

When calculating d, a wider grid than necessary for the imaging would be used, twice as wide is usually suitable. Then, to combine the distances from transducers pairs with different x_{sep}, all that is required is to shift the grid horizontally by x_{sep}, and then to add it to an unshifted version of the grid. This provides the total emitter-scatterer-receiver distance, D. The distances for all the

possible transducer separations are stored in memory; this requires only a small amount of memory unless there is a very large number of transducers or unless the grid is very large. Assuming the same time points for each data collection, these distances can be converted to sample points, n, within each A-scan, relative to a zero-time point, t_0, and sampling frequency, f_s, for a known speed, c:

$$n = \lfloor (d/c - t_0) f_s + 1.5 \rfloor \tag{8.62}$$

$$n = \lfloor (d + (1.5c/f_s - ct_0))(f_s/c) \rfloor \tag{8.63}$$

If there are N time points, values outside of the range $1 \leq n \leq N$ are set to $N+1$, which points at a zero value. The data is loaded for each A-scan, and converted to a fully complex analytic signal. For each emitter-receiver pair, the grid of indexes into the A-scan is shifted horizontally, so that it is centred on the emitter-receiver pair centre, and the imaged region can then be obtained by adding together each relevant data point (n), to each grid element. This summation is performed using the fully complex signal data; the final stage is to take the absolute magnitude of each grid element to convert the complex image to a magnitude image.

These optimisations are pointless if only a single frame is being imaged, but for imaging multiple frames, it means that only the addition of the complex data points needs to be repeated between different frames. All the other calculations need only be performed once.

8.2.13 Basic raycasting

To test all the algorithms within this project, basic simulators were created. These calculate emitter-receiver distances for surface waves and emitter-scatterer-receiver distances for bulk waves. A simulated emission can then be delayed by the appropriate time for the wave speed and distance. Mode-conversion is incorporated by assuming a certain fraction of the incident wave energy is mode-converted at the scattering point. All emitters, receivers, and scatterers, are assumed to be omnidirectional and point-like. The idea of using such a basic simulator to test the algorithms is that for perfect data, the output should be perfect, in terms of it perfectly matching what is expected. If the output is not perfect, the algorithm implementation is flawed, and this is the sort of mistake such simple simulations are designed to find. In addition, these simulations show how the algorithms will perform under ideal conditions, and hence it can be seen if the output is acceptable under such conditions.

Considering first the simulator for TOFDI, required parameters are the initial emitter position, $\mathbf{E_0}$, its movement vector, $\mathbf{u}$, and how many A-scans are to be collected. For each scan, s, the emitter position is calculated:

$$\mathbf{E}(s) = \mathbf{E_0} + s\mathbf{u} \tag{8.64}$$

Receiver offsets are provided as constant positions relative to the initial emitter position, each receiver having a unique offset, $\mathbf{R_n}$. The receivers are then assumed to follow the emitter as it scans across the sample surface, the positions being calculated as:

$$\mathbf{R}(s) = \mathbf{E}(s) + \mathbf{R_n} \tag{8.65}$$

The point-scatterers are defined in terms of an absolute position, $\mathbf{D_n}$, as are the sample edges, $\mathbf{S_n}$, which are assumed perpendicular to the scan direction, and provide reflection points for the surface waves. Required parameters are the speeds, c, of every wave to be simulated; bulk compression and bulk shear waves, lateral compression and Rayleigh waves. Amplitude scale factors are input for each wave mode; these are used to scale the simulated emitted amplitudes. Intrinsic attenuation coefficients, α, are input for each wave mode:

$$A = A_0 e^{-\alpha d} \tag{8.66}$$

A noise level is also requested, and simulated AWGN of this standard deviation is used to corrupt the signal. The sampling rate can be chosen, and the form of the signal can be arbitrary.

The distances to each receiver from each emitter are calculated for the purposes of surface wave arrival by simply taking the magnitude of the receiver offset, $|\mathbf{R_n}|$. Assuming an isotropic and homogeneous medium, the emitter-scatterer distances, r_e, are:

$$r_e = |\mathbf{D_n} - \mathbf{E}(s)| \tag{8.67}$$

The receiver-scatterer distances, r_r, are:

$$r_r = |\mathbf{D_n} - \mathbf{R_n}(s)| \tag{8.68}$$

For both the surface and bulk waves, the distances can be converted to times by dividing by the appropriate speed. It is a mode-converted wave if the emitter-scatterer path and scatterer-receiver path are traversed at different speeds.

The simulated signal is generated (figure 8.7 stretched over twice the time, for a 0.5MHz centre frequency). It is shifted in time by adding the total time of flight, and using linear interpolation. Linear interpolation is suitable here, due to the high sampling frequency leading to negligible errors. For each receiver, the signal is scaled for geometric spreading, with a scaling factor $\propto 1/r_e r_r$, and for intrinsic attenuation, with a scaling factor $\propto e^{-\alpha(r_e+r_r)}$. This is also the case for bulk waves that can be observed at the surface. Surface waves, such as the Rayleigh wave, have a geometric spreading scaling factor $\propto 1/\sqrt{r_e r_r}$. More complicated issues, such as directivity, waves that creep round the defect, rough defects, concave defects, etc. are neglected. Reflections from edges are handled in a similar fashion, essentially treating the edge as a scatterer.

The simulator for the TFM imaging calculates distances, scale factors, etc. in the same way. However, there are multiple static emitters at:

$$\mathbf{E}(s) = \mathbf{E_0} + s\mathbf{u} \tag{8.69}$$

In this case, $\mathbf{E_0}$ represents the start point of the linear array, $\mathbf{u}$ is the array spacing, and s is the element of the array. The receivers are at the same positions.

8.2.14 FEM model creation

Models that are used as the input to the PZFlex FEA software are created using a custom made MATLAB program. The scenario created was a standard TOFD set-up, except that multiple receivers are present. As inputs, the script requires the required simulation time, details of the model parameters such as the damping frequency, number of dimensions (2 or 3), element size

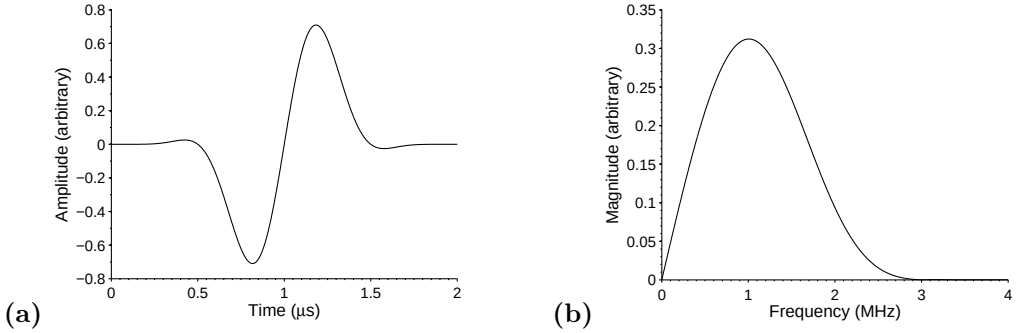


Figure 8.7: The simulated signal input into the simulations, in the time (a) and frequency (b) domains. The signal consists of two complete cycles of a 1MHz sine wave multiplied by a simple window function. The window function has been designed to have a wide main beam but extremely small side lobes. Inputting this signal into the FEA software produces outputs that resemble real signals, and it has the advantage that the frequency domain representation makes it very easy to define a limit for the frequency, above which frequency content is negligible (3MHz in this case).

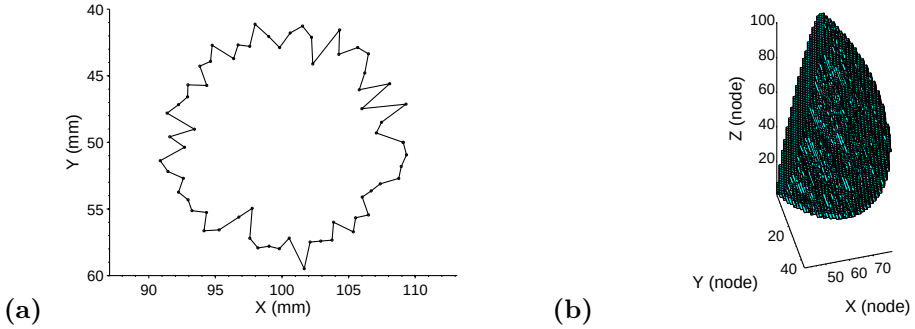


Figure 8.8: Examples of rough (a) and elliptical (b) defects. The rough defect is simply a two-dimensional side-drilled hole of 8mm radius, with AWGN added to the radial extents at a selection of angles. The three-dimensional elliptical defect (depth 5mm, length 10mm, width 0.5mm, vertical angle 30°, horizontal angle 10°) is shown using voxels and is as smooth as the length of an element allows, with this size dictated by other model parameters.

(forced to the nearest sensible increment such as 100µm), details of the generation pulse such as its position, extent, and time domain form, details of the sample such as its size and the form of any defects present, and finally, details of the receiver positions and extent. If information is provided on how the emitter moves, multiple models are produced, one for each emitter position (all receiver positions are captured at each emitter position). The element velocities are output and stored for later analysis. The element velocities are taken rather than the displacement because the FEM software uses velocities as the fundamental quantity it calculates, and then calculates displacements from that by integrating the velocities. This can be done in post-processing if necessary, and therefore the simulation itself need not be slowed by doing so. In addition, EMATs are velocity sensors, and it is therefore this output that is observed by the receiver in practical usage anyway.

The defect specified can be something as simple as a side-drilled hole, or it could be an elliptical crack with variable extents and angles. Roughness of a defect is also supported, such as a rough side-drilled hole. This is modelled as a simple circle in polar coordinates, (ρ, θ), with AWGN added to the radial extent at a selection of angles. Examples of these defects are given in figure 8.8.

Ultrasound is generated by applying a simulated transient pressure load. If this is applied to a single point, both compression and shear waves will have significant energy. If applied to a higher dimensional space (line segment in 2D or an area segment in 3D), then the type of waves generated will depend on the direction of the applied pressure load, but both compression and shear waves will still be generated at the spatial edges of the applied pressure load. The generation pulse (input as a pressure load) used is shown in figure 8.7. Note that this generation pulse is not designed to exactly represent what the transducer produces experimentally. Instead, it is designed to be easy to model, with a variable centre frequency, and it produces pulses that when received in the model, approximate pulses received in the experimental case. The pressure load was applied either perpendicular or parallel to the surface; perpendicular to the surface is a better approximation to an ablative laser source, and was therefore generally favoured for these tests. If the applied pressure load is perpendicular to the surface, compression waves will be generated primarily. If the applied pressure load is parallel to the surface, shear waves will be generated primarily. Either load will generate both compression and shear waves to some extent.

The 100μs simulation time, the 100μm length of an element, and the steel material characteristics, were kept constant. Steel was defined as a viscoelastic material (a viscoelastic material dissipates energy as heat), with a density of 7900kg/m^3, a compression wave speed of 5960m/s that has an 0.3dB/cm amplitude attenuation at 1MHz, and a shear wave speed of 3235m/s that has a 1.2dB/cm amplitude attenuation at 1MHz. The damping follows an exponential decay. The pressure load was modelled as a perfect line source with a width of 200μm but a much larger (or even infinite) length. All sides of the model were absorbing boundaries, except for the top side which represents the sample surface, and upon which the signal was generated; this was a free boundary. An absorbing boundary approximates an infinite extent past that point. The defect material was always taken to be a perfect vacuum. Using air made an insignificant contribution to the ultrasound incident at the realistic receiver positions, and hence complicated the model unnecessarily.

The particular defects modelled were: smooth side-drilled holes, with radii of 1mm, 2mm, 4mm, and 8mm; smooth elliptical holes, with a semi-minor axis of 1mm, and semi-major axes of 2mm, 4mm and 8mm (orientations both horizontal and vertical); the circular holes again, but with added roughness. These were in the centre of a two-dimensional steel sample (and hence are extended infinitely in the third dimension) of depth 10cm and length 20cm.

For these models, numerical dispersion due to discretisation was not believed to be a significant issue, and is not addressed further here. It was confirmed that the models had converged, by halving the number of elements used in the simulation, and confirming the output was the same to a good tolerance.

8.3 Simulated results

8.3.1 Raycasting

The setup is depicted in figure 8.9. The TFM array consisted of 32 elements with a 1mm spacing, with the first element at $(-15.5\text{mm}, 0)$, and extending along the x-axis to $(15.5\text{mm}, 0)$. The TOFDI emitter scanned from $(-95.5\text{mm}, 0)$ to $(95.5\text{mm}, 0)$, with a 1mm spacing, and the two receivers were placed 80mm either side of the emitter, producing B-scans as shown in figure 8.10. In both cases, the signal centre frequency was 500kHz, and the sample was considered an infinite half-space.

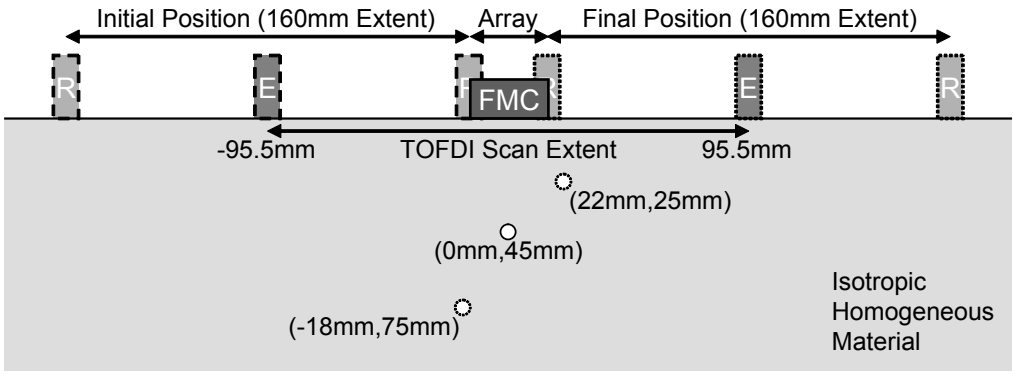


Figure 8.9: Simulation geometry for the raycasting results. The TFM array (performing FMC) consisted of 32 elements with a 1mm spacing, with the first element at $(-15.5\text{mm}, 0)$, and extending along the x-axis to $(15.5\text{mm}, 0)$. The TOFDI emitter (E) scanned from $(-95.5\text{mm}, 0)$ to $(95.5\text{mm}, 0)$, with a 1mm spacing, and the two receivers (R) were placed 80mm either side of the emitter. A point scatterer was placed at $(0, 45\text{mm})$, and for later tests, two additional point scatterers were placed at $(22\text{mm}, 25\text{mm})$ and $(-18\text{mm}, 75\text{mm})$. The resultant B-scans are input into TOFDI and TFM algorithms.

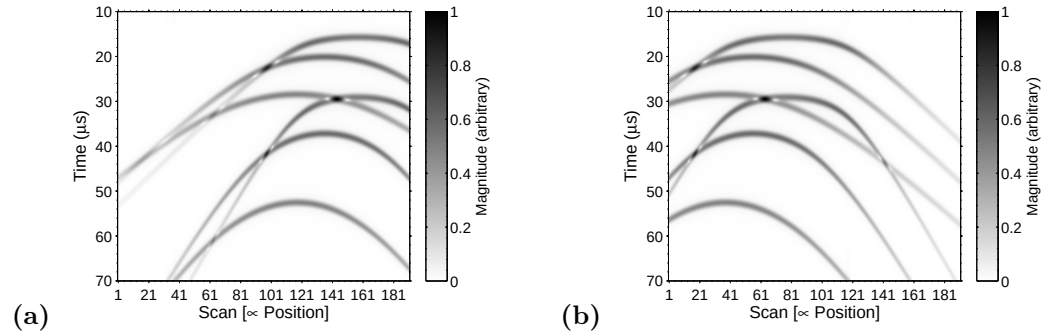


Figure 8.10: Using the setup in figure 8.9, with all three defects, these are the B-scans input into the TOFDI algorithm for the left (a) and right (b) receivers. The complex analytic signal magnitude is shown here.

The waves speeds are 5960m/s for compression waves, 3235m/s for shear waves, and 2996m/s for Rayleigh waves. The simulation is basic and idealised; it includes compression and shear waves, but no mode conversions or surface waves. The TFM system only handles compression waves, but shear waves are included in the data set such that a second defect may be observed due to the shear waves being treated as compression waves.

For the first test, the point scatterer was placed at $(0, 45\text{mm})$. Putting the B-scans into TOFDI (summing the images from each receiver to produce the overall cross-sectional image) and TFM imaging algorithms, produces figure 8.11, using the complex analytic signal. The alternatives are using the real data without the analytic transform (figure 8.12), using the magnitude of the analytic signal (figure 8.13), and using the phase of the analytic signal (figure 8.14). As expected, using the full analytic signal, with magnitude and phase components, produces the best image in terms of point spread and image artefacts. Consequently, for all further analysis, the fully complex analytic signal will always be used in preference to the alternatives. The full width at half maximum (FWHM) of the TOFDI peak is 5.8mm laterally, and 3.8mm in depth, and the FWHM of the TFM peak is 14mm laterally and 4.2mm in depth. The smaller point spread of the TOFDI imaging is only due to the far larger lateral region scanned. If the TFM array existed over

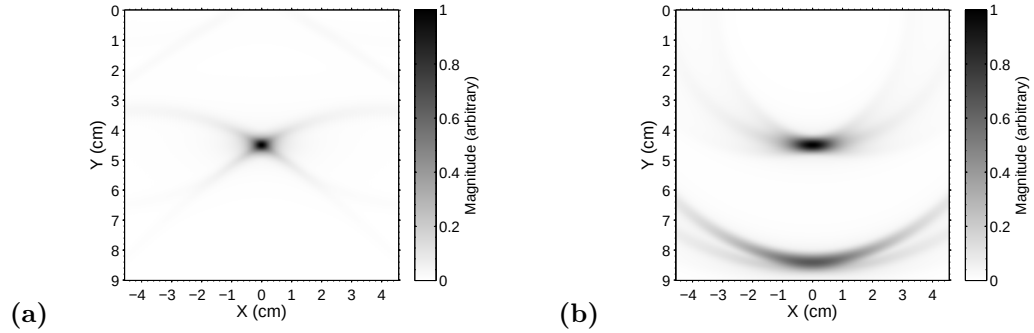


Figure 8.11: Using the setup in figure 8.9, with only the first defect, this imaging used the signal after conversion to a complex analytic signal. TOFDI (a) and TFM (b) images have good resolution.

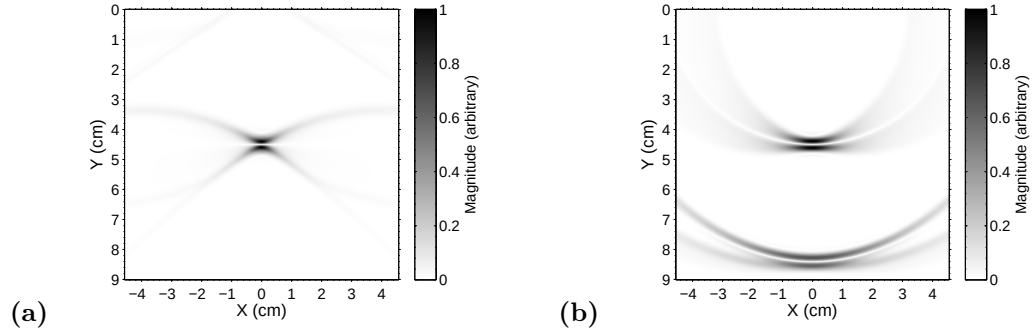


Figure 8.12: Using the setup in figure 8.9, with only the first defect, this imaging used the real signal without any conversion. TOFDI (a) and TFM (b) images still have good resolution, but now the form of the signal can be seen in the imaging output. In particular, the central trough has occurred at the scatterer position. As this would complicate image interpretation, using the real signal cannot be advised over using the analytic signal.

a similar region, the lateral point spread would at least match and probably exceed that of the TOFDI image. This point spread will vary depending on the position of the defect relative to the TFM array and the TOFDI array scan, in terms of both lateral position and depth. The defect position is representative of the sort of defect position that is being examined experimentally. The TFM image has a false defect at the bottom due to treating the shear waves as compression waves. The TOFDI image would have a similar false defect, but due to the much larger emitter-defect-receiver distances, the false defect appears at a greater depth than is included in the imaged area. If the TFM array existed over a similar region to the scan of the TOFDI array, this false defect would very likely have such a small magnitude that it would effectively not exist.

For the second test, two additional point scatterers were placed at $(22\text{mm}, 25\text{mm})$ and $(-18\text{mm}, 75\text{mm})$. Putting the B-scans, after conversion to a complex analytic signal, into the TOFDI and TFM imaging algorithms, produces figure 8.15. Considering first the TOFDI image, the lateral point spread varies slightly with deviation from the centre of the scan, and depth point spread clearly varies with depth, with shallower defects having greater depth spread relative to deeper defects. The point spread of the three points in figure 8.15a, left to right, are $(6.5\text{mm}, 3.4\text{mm})$, $(5.6\text{mm}, 3.9\text{mm})$, and $(6.6\text{mm}, 5.0\text{mm})$. There are two false defects, but at a much lower magnitude. The false defect at the top-left, is due to the shear wave of the bottom-left defect being imaged as a compression wave, in addition to being treated as a shear wave. The false

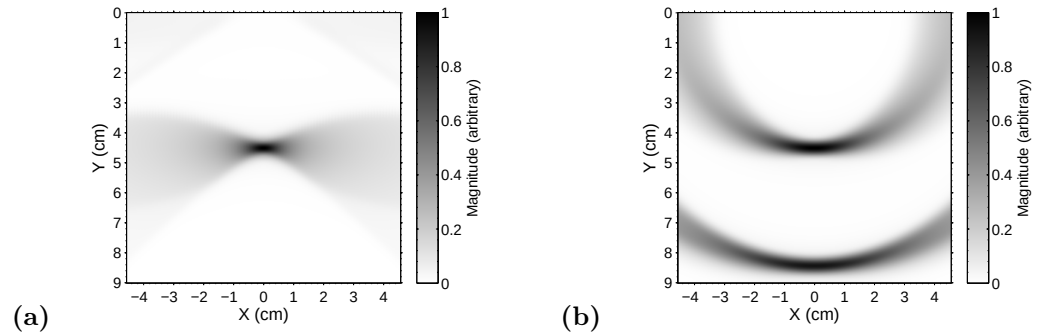


Figure 8.13: Using the setup in figure 8.9, with only the first defect, this imaging used the magnitude of the analytic signal. TOFDI (a) and TFM (b) images have lost some resolution, and image lobes are more pronounced. This is because the signal has amplitude in these regions, but when using the analytic signal, the different phases cancel to minimise the lobes; using the magnitude, these parts of the signal can only combine constructively. Due to this, using the magnitude of the analytic signal cannot be advised over using the fully complex analytic signal.

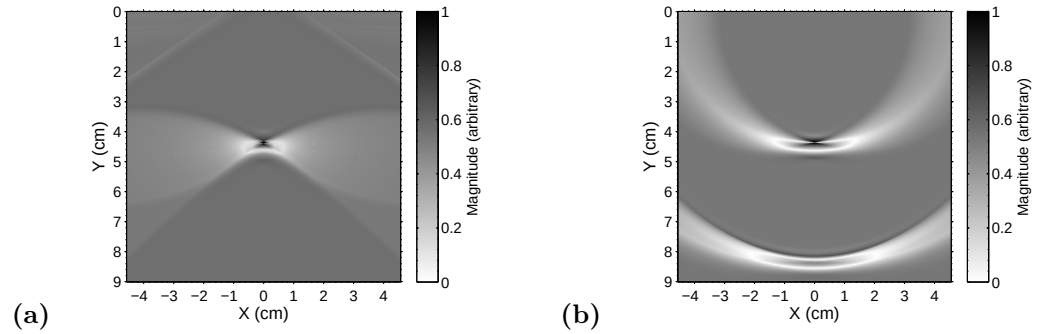


Figure 8.14: Using the setup in figure 8.9, with only the first defect, this imaging used the phase of the analytic signal. TOFDI (a) and TFM (b) images have lost some resolution, and image lobes are more pronounced. In addition, the signal is less clear. When using the full complex analytic signal, the lower magnitude regions contribute less to the image than the high magnitude regions, and low magnitude regions are more easily cancelled out by the interference of high magnitude regions. However, using just the phase removes the natural weighting from the magnitude component of the signal. Due to this, using the phase of the analytic signal cannot be advised over using the fully complex analytic signal.

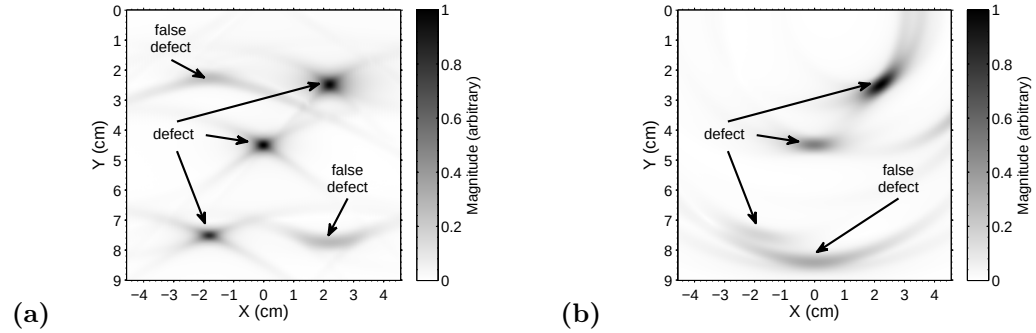


Figure 8.15: Using the setup in figure 8.9, with all three defects, this imaging used the fully complex analytic signal. TOFDI (a) and TFM (b) images both have false defect indications and point spreads that vary with defect position. Smaller depth point spread is achieved at greater depths, and better lateral point spread is found at the centre of the array (or array scan). No distance compensation has been applied in either case.

defect at the bottom-right, is due to the compression wave of the top-right defect being imaged as a shear wave, in addition to being treated as a compression wave. Considering the TFM image, there is a false defect at the bottom-middle, due to the shear wave of the central defect being treated as a compression wave. Of more interest is the variable point spread. The three indications are tilted at different angles (the centre one being flat), due to the offset from the lateral centre of the array. A similar thing is not seen with the TOFDI image due to the large lateral scan. As before, if the TFM array existed over a similar region to the TOFDI scan, the point spread would at least match, and probably exceed, that of the TOFDI image. The tilting only occurs due to the limited size of the TFM aperture. The tilting would also be removed by laterally moving the TFM array such that it was over each defect, and combining the images output from each position (assuming the position of the array was being tracked). The defect indications in the TFM case are also of very different magnitude; no distance compensation is applied during either imaging process, and so it is expected that deeper defects will not appear as strongly. This is less pronounced in the TOFDI case, but only because the large lateral emitter-defect-receiver distance means that changes in depth are relatively much smaller. Application of distance compensation would solve the issue, but this would also enhance the indication of the false defect. Again, false indications could be reduced by a larger TFM aperture, or by combining TFM images when the TFM array is at different positions. The defect highlighting stage of the TOFDI algorithm correctly identified the position of all three defects, as would be expected for such an ideal input.

It has been stated that one of the advantages of the cross-sectional imaging technique is that the output is considerably easier to interpret than the B-scans used as input. Comparing figures 8.10 and 8.15, even without surface waves (received directly and reflected off edges), back-wall reflections, and mode-converted waves, the presence of both compression and shear waves makes the interpretation of the B-scans complicated; it would certainly require much more training than the interpretation of the cross-sectional image. The overlapping parabolas will confuse distance compensation, and this is part of the reason why such compensation has not been applied thus far.

All of the tests performed so far, were repeated using the alternative TOFDI technique of taking the product of the images from the two receivers. This produces figure 8.16, figure 8.17, and figure 8.18. Again, using the full analytic signal, with magnitude and phase components, produces the best image in terms of point spread and image artefacts, as can be seen by comparing figure 8.16a

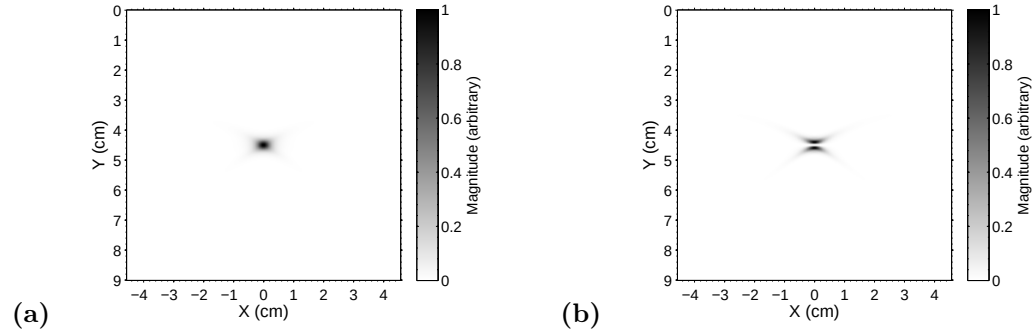


Figure 8.16: Using the setup in figure 8.9, with only the first defect, these TOFDI images used the fully complex analytic signal (a) and the real signal without any conversion. The product of the images generated by each receiver has been used rather than the summation.

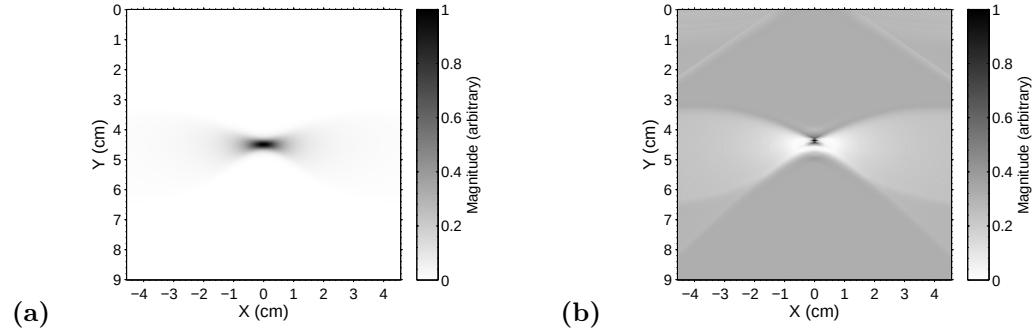


Figure 8.17: Using the setup in figure 8.9, with only the first defect, these TOFDI images used the magnitude of the analytic signal (a) and the phase of the analytic signal (b). The product of the images generated by each receiver has been used rather than the summation.

(analytic signal input), against 8.16b (real signal input), 8.17a (analytic signal phase input), and 8.17b (analytic signal magnitude input). From figure 8.16a, the FWHM of the TOFDI peak is 3.8mm laterally, and 2.6mm in depth. Not only has the point spread improved, image artefacts have also been minimised, as is particularly noticeable in figure 8.18. However, in this case, the improvement is not real. Due to the simplicity of the simulator, both the receivers detect the exact same signal during the scan; it is merely shifted in scan position. Consequently, taking the product of the cross-sectional images produced by each receiver (figure 8.18a) is identical to taking the square of the summed images (figure 8.18b). This would suggest that no improvement can be achieved by taking the product, but this conclusion cannot be drawn due to the simplicity of the simulator and the symmetry of the receivers either side of the emitter. In a more realistic simulator or a non-symmetrical case, the two receivers will not detect the exact same signal, and taking the product might still show improvement over taking the sum and then squaring the result. Indeed, in this simplistic case, there is not even an advantage to having two receivers over having one, if both receivers are at the same separation (but different sides).

8.3.2 Finite element analysis

The FEA software PZFlex was used to generate an ideal case that was more realistic than the basic simulator. The setup is depicted in figure 8.19. The TFM array consisted of 32 elements with a 1mm spacing, with the first element at $(84\text{mm}, 0)$, and extending along the x-axis to $(115\text{mm}, 0)$;

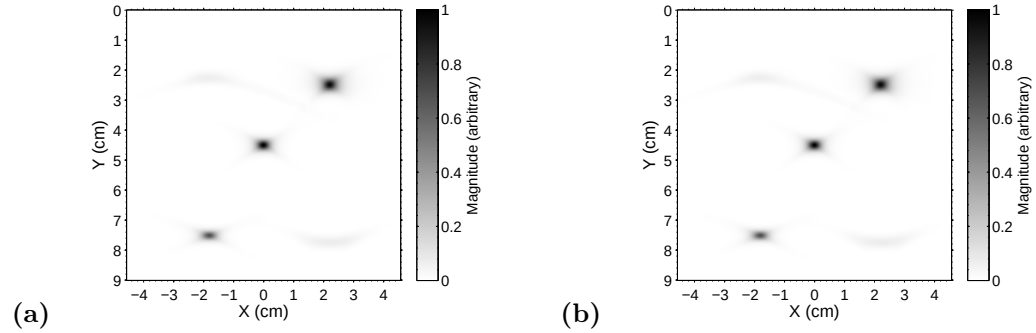


Figure 8.18: Using the setup in figure 8.9, with all three defects, these TOFDI images used the fully complex analytic signal. Compared here, is the use of the product of the images generated by each receiver (a), and the square of the summation of the images generated by each receiver (b). As the two input B-scans are identical except for a shift in scan position, the two images are identical. The basic nature of this simulator is unable to show any improvement due to taking the product rather than the square of the summation.

this actually means the array centre is just off the sample and defect horizontal centre. The TOFDI emitter scanned from $(10\text{mm}, 0)$ to $(190\text{mm}, 0)$, with a 1mm spacing, for 181 different emitter positions. The pressure load was applied perpendicular to the surface (parallel to the depth), as this is the force direction for an ablative laser source. Two receivers were placed 80mm either side of the emitter. Unfortunately, data was not collected for all receiver positions; when a receiver is between $(10\text{mm}, 0)$ and $(190\text{mm}, 0)$, the simulated lateral extent that has reliable output, data is used from the simulation (at the same 181 points that the emitter fired at), but outside of this region, the data is replaced with zeros. This can be seen in figure 8.20, the B-scan for a 2D 2mm circular defect at the centre of the sample $(100\text{mm}, 50\text{mm})$. The size of the two-dimensional steel sample, of depth 100mm and length 200mm, was limited due to time and computer memory constraints.

The TFM system only uses compression waves, but both TOFDI and TFM simulations were receiving compression, shear, and mode-converted waves. Recorded were the x-direction velocities; EMATs are velocity sensors, and the ones used within this project were primarily sensitive to motion parallel to the surface. Velocities can be converted to displacements by integrating with respect to time, but this has not been performed, as it only changes the shape of the waveform, not when it arrives or its duration, and hence has no real impact on the imaging performance observed. Whenever possible, surface waves and other signals not from the scatterer are eliminated using a background subtraction. The PZFlex model used for a background had no scatterer in the model, so waves observed were directly received surface waves and back/side-wall reflections.

Based on the earlier analysis of the basic raycasting model, the full analytic signal, with magnitude and phase components, will be used in the imaging process. This produces the best image in terms of point spread and image artefacts. The reason this works is that the signal from a point scatterer is expected to be phase coherent (have a constant phase) as the emitter-receiver pair scans over it. Consequently, the parts of the signal that are scattered from the defect, and represent the same time point in the signal pulse, combine coherently (as they have the same or very similar phase), and hence become larger, whereas other signals, or noise, do not combine coherently, and hence become smaller. To check the signal is phase coherent, the signal from just the defect is isolated (figure 8.21a), such that each identical time point of the signal can be easily

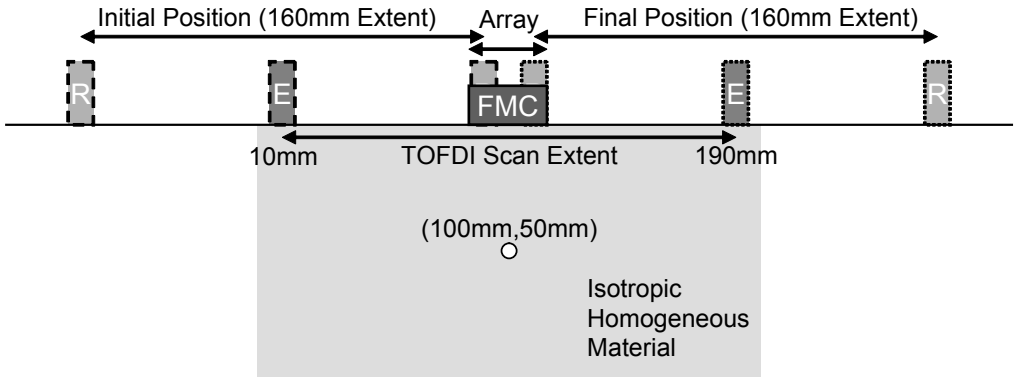


Figure 8.19: Simulation geometry for the FEA results. The TFM array (performing FMC) consisted of 32 elements with a 1mm spacing, with the first element at $(84\text{mm}, 0)$, and extending along the x-axis to $(115\text{mm}, 0)$. The TOFDI emitter (E) scanned from $(10\text{mm}, 0)$ to $(190\text{mm}, 0)$, with a 1mm spacing, and the two receivers (R) were placed 80mm either side of the emitter. The defect (various forms were used) was centred at $(100\text{mm}, 50\text{mm})$. The resultant B-scans are input into TOFDI and TFM algorithms. The simulated region (the grey region of isotropic homogeneous material) was a steel sample of depth 100mm and length 200mm. At the TOFDI scan extremes, data was not collected for all TOFDI receiver positions, as the receivers were beyond the simulated region.

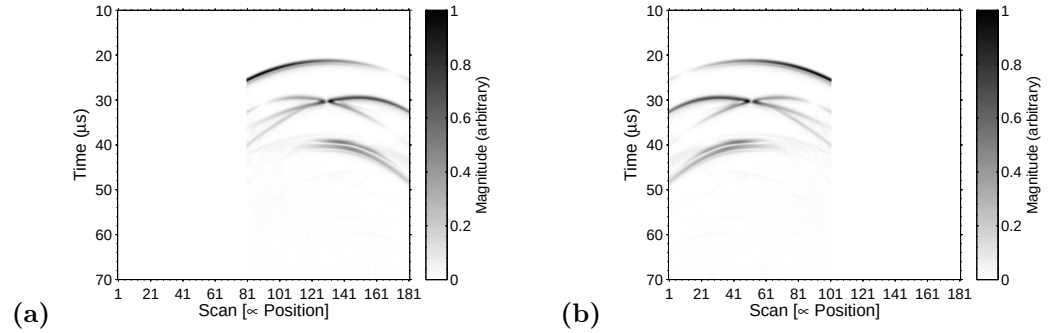


Figure 8.20: Using the setup in figure 8.19, with a 2mm diameter side-drilled hole centred at $(100\text{mm}, 50\text{mm})$, these are the B-scans input into the TOFDI algorithm for the left (a) and right (b) receivers. The complex analytic signal magnitude is shown here. As expected, scan (b) is simply scan (a) flipped along the vertical axis, but the scans are not symmetrical about the point for which the defect is the same distance from the emitter and receiver. Consequently, it is not expected that taking the product and the square of the summation will produce the same result.

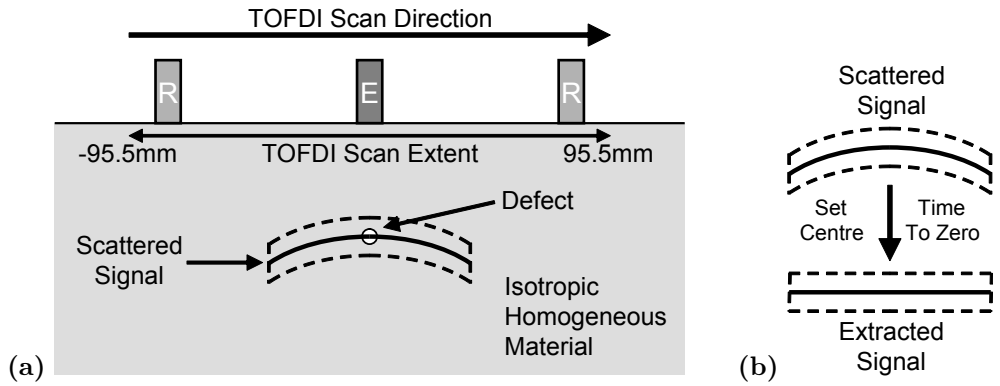


Figure 8.21: The phase coherence of scattered signals is important to the imaging process if using the fully complex analytic signal. To check the phase coherence, the signal in the B-scan representing the ultrasound scattered from the defect is isolated. Here, the parabola passing through the defect represents the scattered signal, and the dashed lines around it represent the isolated part of the B-scan (a). The centre of the extracted region, again represented by the parabola passing through the defect, is set as the zero time point (b). This makes all the sections of extracted signal line up in time, such that the centre of the signal pulse is at the same time point. It is then simple to obtain the analytic signal, and hence to extract the phase (figure 8.22).

compared (figure 8.21b). The phase is compared for the isolated signals, and phase coherence is indeed observed in the simulated data (figure 8.22). Note that the change in phase after A-scan 171 (figure 8.22a) and before A-scan 11 (figure 8.22b) is because the receiver is passing over the scatterer, and the situation is transferring from one in which the emitter and receiver are either side of the scatterer, to one in which the emitter and receiver are the same side of the scatterer. Due to this change in phase coherence, there is no benefit in having the pair boundary for a given emitter-receiver pair extending beyond the horizontal separation of the emitter and receiver. In addition, past this point, the magnitude of the signal is much weaker than when the defect is between the emitter and receiver.

Tested earlier was the TOFDI technique of taking the product of the images from the two receivers. This appeared to be equivalent to taking the square of the summation of the images from the two receivers. Figure 8.23a is the result of taking the summation, and 8.23b is the result after squaring the summation. Comparing this to the result of using the product instead, figure 8.24b, the improvement from using the product is minimal, because the defect signal is so much stronger than the artefacts. For the experimental data, the artefacts are stronger than the signal, even after summing the signals from the two receivers; under these circumstances, there is a considerable advantage to using the product of the signals from the two receivers rather than the square of the summation. This is because the defect position remains the same in both images, whereas the position of artefacts varies; taking the product enhances the constant defect position much more strongly than summation, and therefore relatively reduces the shifting artefacts. Consequently, the product will be used in preference for all future TOFDI cross-sectional images.

If only a single wave mode (compression or shear) is present, then only that needs to be considered in the TOFDI imaging process. However, more generally, compression and shear waves will be present, and mode-conversions between the two. Under certain circumstances, it is better to only consider a single wave mode despite the presence of multiple wave modes. For example, figure 8.25 considers just using compression or shear waves. Only considering shear waves in figure 8.25b has resulted in strong image artefacts due to the compression wave and the mode-converted waves

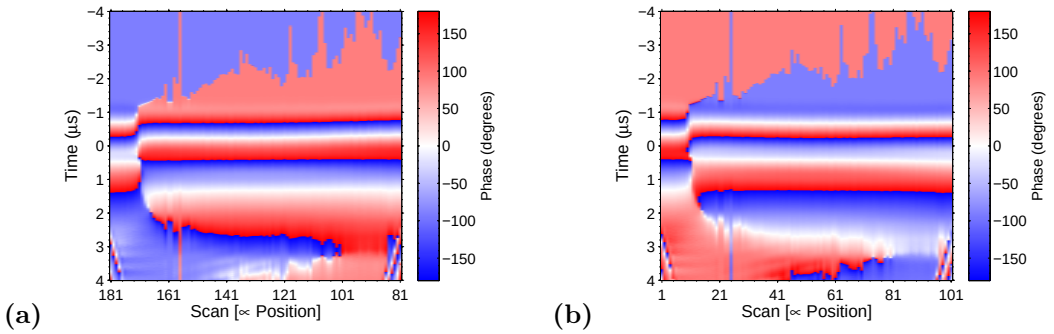


Figure 8.22: The phase coherence of scattered signals can be investigated by extracting from figure 8.20 the signal scattered by the 2mm diameter side-drilled hole, centred at (100mm, 50mm), using the method described in figure 8.21. Shown here is the phase of the analytic signal extracted from the B-scan. This signal has been displayed such that the time corresponding to the emitter-scatterer-receiver path is represented by zero time, hence changing the arrival times from a parabola to a horizontal line, as in figure 8.21b. Receiver 1 (a) and 2 (b), representing the left and right receivers respectively in figure 8.21a, both display phase coherence for most of the scan. The scan axis of (a) has been reversed for an easier comparison with (b).

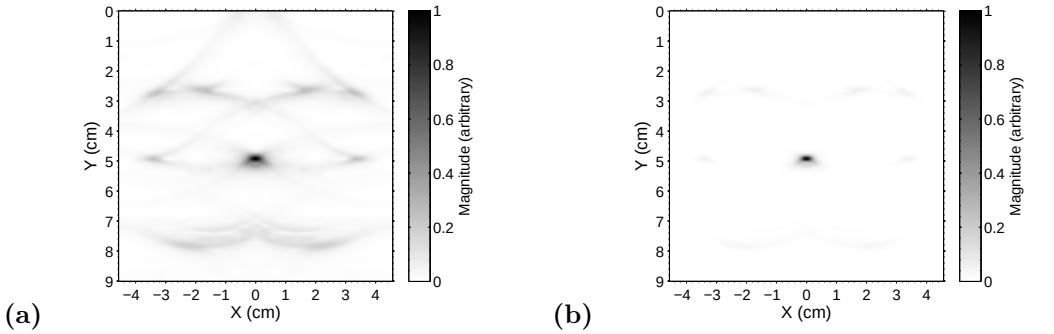


Figure 8.23: Using the setup in figure 8.19, with a 2mm diameter side-drilled hole centred at (0cm, 5cm) as it appears in the images, this imaging used the fully complex analytic signal. Compared here is use of the summation of the images generated by each receiver (a) and the square of the summation of the images generated by each receiver (b).

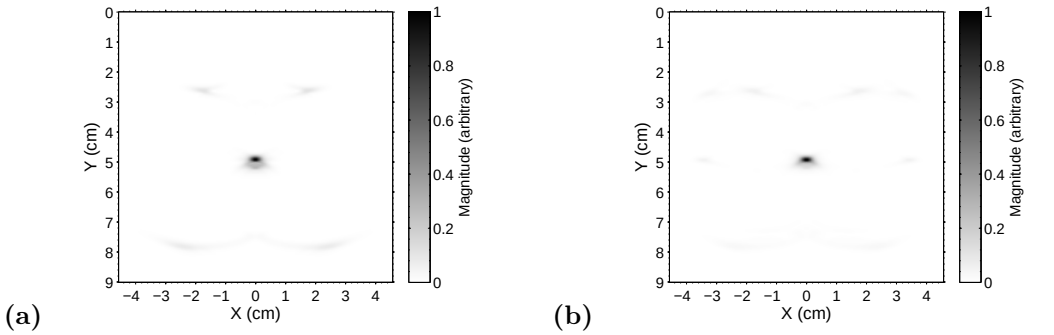


Figure 8.24: Using the setup in figure 8.19, with a 2mm diameter side-drilled hole centred at (0cm, 5cm) as it appears in the images, this imaging compares the output from treating all the received signals as compression or shear waves (a), or as compression or shear waves or waves that have mode-converted (b). The images are very similar.

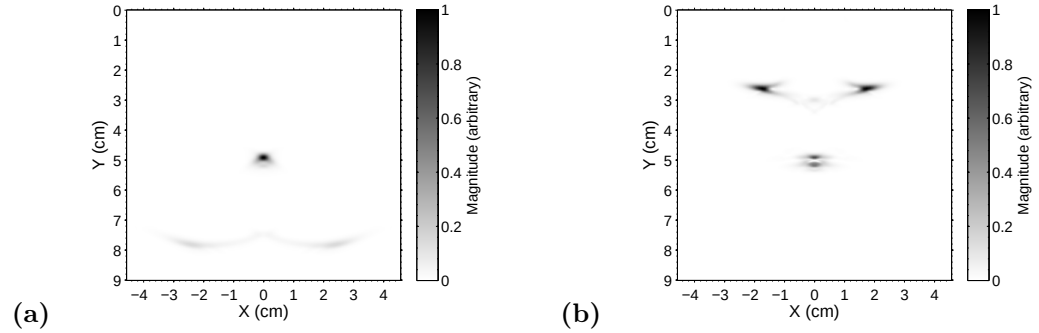


Figure 8.25: Using the setup in figure 8.19, with a 2mm diameter side-drilled hole centred at (0cm, 5cm) as it appears in the images, this imaging compares the output from treating all the received signals as compression waves (a), or as shear waves (b). Image (b) has considerably worse image artefacts than image (a), but this is partly because the strong image artefacts for (a) appear below the bottom of the imaged area. This property cannot be relied upon generally.

also being treated as shear waves. Such strong image artefacts are not present in figure 8.25a, which only considers compression waves. In this case, the weak image artefacts are due to the mode-converted waves being treated as compression waves, but the strong image artefacts, due to treating the shear waves as compression waves, is below the bottom of the imaged area. However, in general, strong image artefacts being below the imaged area cannot be relied upon. Figure 8.24a uses compression and shear waves, and figure 8.24b uses compression and shear waves and mode-conversions between them. In this case, figure 8.24b provides the sharpest defect indication and the least image artefacts. Mode-converted waves are not observed at every angle, and potentially the imaging could be improved by incorporating the directionality of the mode-converted waves, but this has not been tested at this stage due to insufficient information on such directionality. It is also possible that in the case of a general rough defect, such directionality may be variable. The range of possible locations of the defects may result in a single wave mode being interfered with by surface waves. Although attempts are made to subtract surface waves in a way that leaves the scattered waves intact within the B-scan, it is still better to have some minor image artefacts, than to miss a defect due to not processing a wave type, and hence all wave modes will be used to form the images, unless it is known a certain wave mode is missing. Adding additional receivers at different separations will reduce the level of the image artefacts relative to the true defect indications.

Having decided on the operational details of TOFDI, its performance against various simulated defects can now be tested. Results have also been obtained for the simulated TFM system previously described. The results shown here are for defects which are at the centre of the sample, (0mm, 50mm) in the cross-sectional images. These defects include a smooth side-drilled hole with a radius of 1mm (figure 8.26), a smooth vertically orientated elliptical hole with a semi-minor axis of 1mm and semi-major axes of 4mm (figure 8.27), and a circular hole with a radius of 8mm and added roughness (figure 8.28). The 1mm radius circle is often used as an example point scatterer when performing experiments, and as simulated here, it does return an indication at the correct point for both TOFDI and TFM in figures 8.26a and 8.26b respectively. The TOFDI algorithm has very little in the way of imaging artefacts, whereas the TFM algorithm produces artefacts, due to signals being present that are of a different wave mode to the one being used for that part of the imaging process. The vertically orientated elliptical defect is used to represent a bulk defect of the

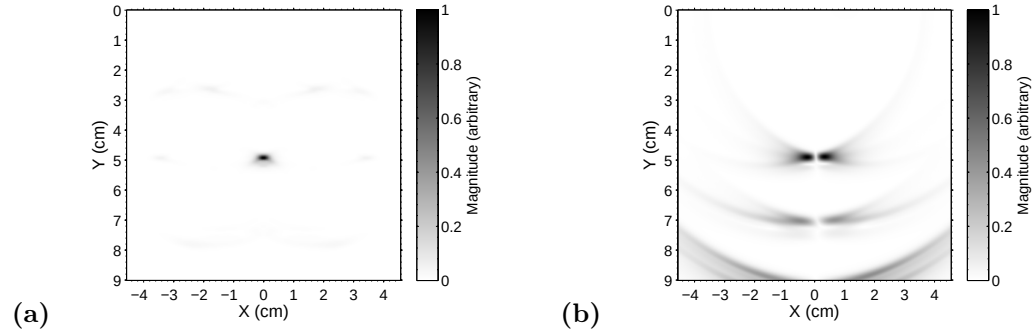


Figure 8.26: Compared here is the output from TOFDI (a) and TFM (b), for a circular 2D defect at (0cm, 5cm) as it appears in the images, of 2mm diameter. Both images have an indication at the correct position.

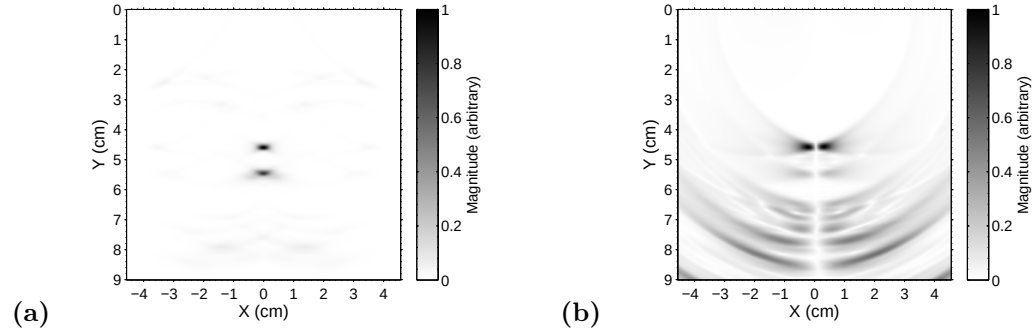


Figure 8.27: Compared here is the output from TOFDI (a) and TFM (b), for an elliptical 2D defect at (0cm, 5cm) as it appears in the images, of 8mm vertical length and 2mm horizontal width. Both images have an indication at the top and bottom of the defect.

sort conventional TOFD would be expected to detect and size (along the long vertical dimension), and as can be seen in figure 8.27a, the top and bottom is visible in the TOFDI image, and could be sized (in terms of vertical length) from this image. The automatic peak finder algorithm outputs the top and bottom of the defect accurately, although the bottom does appear 0.04mm too deep, for the reasons discussed in section 8.2.5. The TFM image in figure 8.27b has not clearly shown the bottom of the defect due to the proximity of the strong nearby image artefacts. This has occurred due to the limited lateral view afforded by the small, static array. The rough defect of figure 8.28 is meant to illustrate the sort of signal that could be expected from a larger volumetric defect; the simulated shape is that in figure 8.8a. In the TOFDI case, the top of the defect is clearly visible, but the bottom is difficult to see, so sizing would not be possible. It is not clear how this limitation could be alleviated. In the TFM case (figure 8.28), the top and bottom of the defect is visible, but sizing would be very difficult as the bottom is accompanied by image artefacts in the spatial vicinity. Investigating the B-scan directly for this case, figure 8.29b, it is not clear that there is a signal from the bottom for the TOFDI imaging algorithm to process, and the roughness makes the scan more difficult to interpret than the equivalent smooth defect case of figure 8.29a.

In all the cases discussed here, if the TFM array was made larger to match the extent of the TOFDI scan, or if scanning was incorporated into TFM in such a way that the results could be combined into a single image, the images output would at least match, and likely surpass, the TOFDI results. The limitation has been the small size of the TFM array aperture, and the point

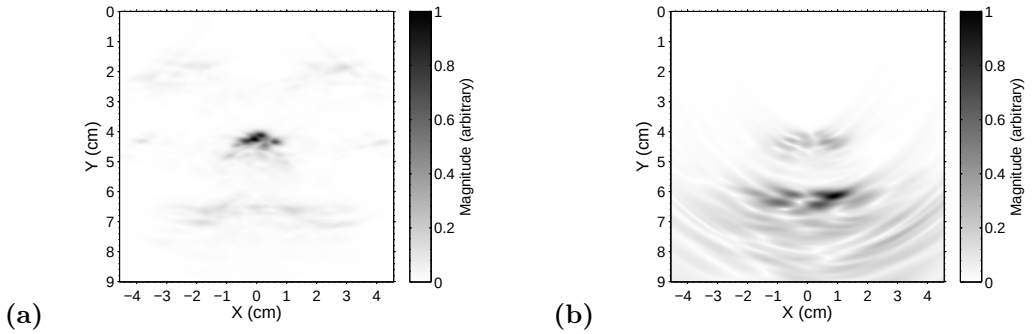


Figure 8.28: Compared here is the output from TOFDI (a) and TFM (b), for a circular 2D defect at $(0\text{cm}, 5\text{cm})$ as it appears in the images, of 2mm diameter, with roughness added (effectively noise in the perimeter of the defect). Both images have an indication at the correct positions, but they are not well defined.

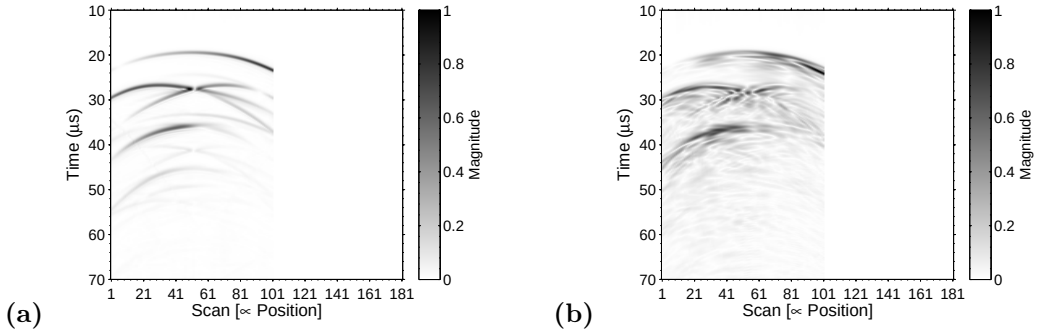


Figure 8.29: Using the setup in figure 8.19, and considering just the right receiver, these are the B-scans input into the TOFDI algorithm for a smooth (a) 16mm diameter circular defect centred at $(100\text{mm}, 50\text{mm})$, and the rough equivalent (b). The roughness makes the B-scan more difficult to interpret than a standard B-scan for a smooth defect.

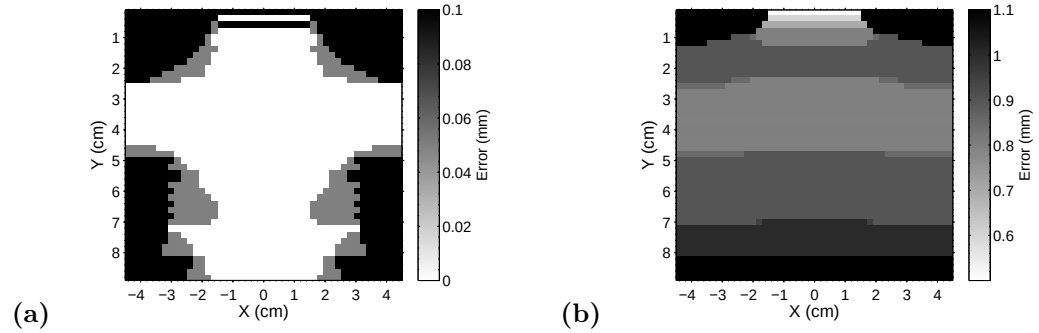


Figure 8.30: By varying the assumed wave speed used in the cross-sectional imaging algorithm, the errors due to an inexactly known wave speed can be assessed. In this case, the error is a ±1% error in shear wave speed. Using the setup in figure 8.9, but with a point scatterer placed at a series of points within the imaged area, the errors in the assessed x-position (a) and y-position (b) are shown for each possible scatterer position.

has been to demonstrate that TOFDI can provide a solution in cases where scanning a TFM array, or having a much larger TFM array, are not practical.

8.3.3 Errors

Errors have been calculated using a very simple method. Using the basic raycasting simulation, ideal simulated data is created for points within the imaged area, neglecting surface waves, and only considering waves of one speed. For each point, the position of the scatterer is assessed using the cross-sectional imaging algorithm, again focusing on only one wave speed. A parameter used in the imaging process, the wave speed the imaging assumes for example, is then varied by a certain percentage in the imaging algorithm (but maintained at its original value in the simulation), and the difference in the calculated position measured. This difference is used as a measure of the error in the scatterer position due to an error in the measured value of that parameter. Figure 8.30 shows the error due to a ±1% error in shear wave speed, figure 8.31 shows the error for a ±5% error in shear wave speed, and figure 8.32 shows the regions that were not correctly imaged even when the wave speed was known exactly. Note that other than the scatterer position, the simulation is created using the same parameters as for the prior results shown, and the imaging algorithm uses the same parameters as for the prior raycasting results shown, with the exception of the wave speed.

The errors can be calculated for whatever parameters are of interest, with whatever scale errors are of interest. Errors for different parameters can be combined using the functional approach [190]. The largest error is expected to be in the measurement of the wave speed, hence it has been the focus of attention here.

8.4 Experimental results

Experimentally, all results are for TOFDI; there is no comparison with TFM as suitable equipment was not available. The experimental B-scans are generated using a TOFD type setup (figure 7.1a). The two linearly wound EMAT receivers are situated either side of the ablative laser ultrasound source, with a separation $s = 7.8$cm, a shear wave speed of $c_s = 3226$m/s, and pulse period of $T = 1\mu$s (approximately). Although this setup can generate and receive multiple wave modes,

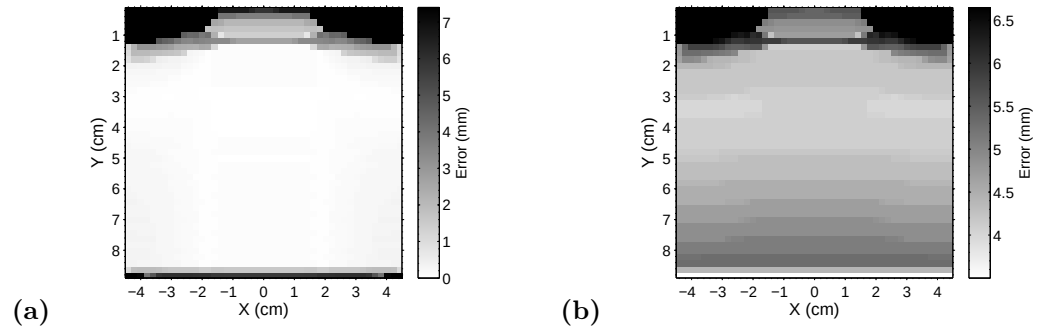


Figure 8.31: By varying the assumed wave speed used in the cross-sectional imaging algorithm, the errors due to an inexactly known wave speed can be assessed. In this case, the error is a $\pm 5\%$ error in shear wave speed. Using the setup in figure 8.9, but with a point scatterer placed at a series of points within the imaged area, the errors in the assessed x-position (a) and y-position (b) are shown for each possible scatterer position. Comparing (a) here with figure 8.30a, due to the range of the scale being much larger here, the errors at the sides of the image appear smaller, when they are at least as large as the errors seen in figure 8.30a.

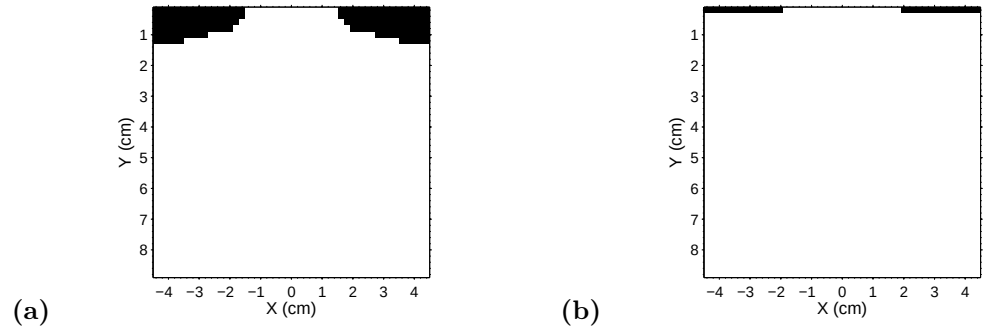


Figure 8.32: Using the same setup as described in figure 8.30 and figure 8.31, the regions that were not correctly imaged even when the wave speed was known exactly are shown as black regions for incorrect x-positions (a) and y-positions (b). These regions will contract if more extensive scans are performed.

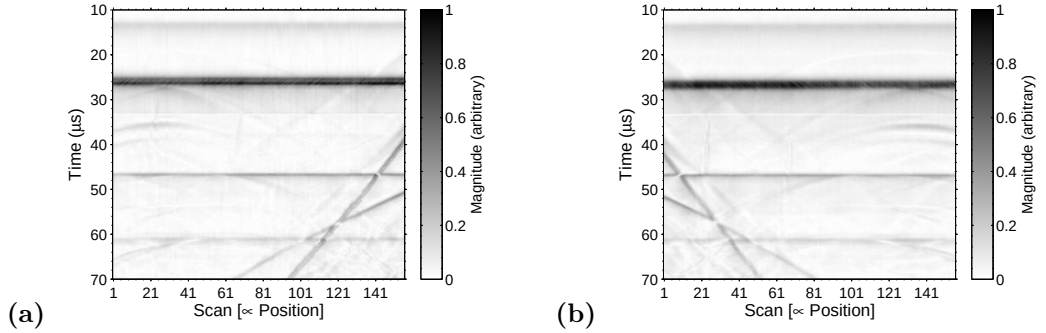


Figure 8.33: B-scans collected during the data capture stage of TOFDI, by the ablative laser ultrasound source and the two EMAT receivers, with magnitude as the luminance. Each scan increment corresponds to 0.8mm of motion, leading to a total scan motion of 124mm. The mild steel sample had a 6mm diameter side-drilled hole at a depth of 45mm, in a position such that the emitter was directly above it for scan 78, placing it directly at the centre of the sample. Due to the symmetry, scan (b) is similar to scan (a) flipped along the vertical axis, and as with the simulated case (figure 8.20), the scans are not symmetrical about the point for which the defect is the same distance from emitter and receiver.

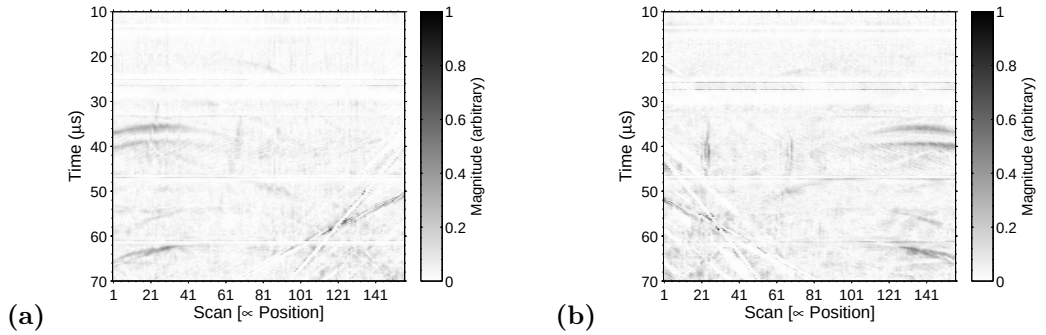


Figure 8.34: Automatically processing figure 8.33 to remove horizontal and angled linear features, converts figure 8.33a to (a) and figure 8.33b to (b). The shear waves from the top and bottom of the defect can now clearly be seen, and the input is ready to be processed by the cross-sectional imaging stage.

including compression and shear waves and conversions between the two, only shear waves (without mode conversion) are clearly present in the B-scans shown here (figure 8.33). The compression waves have insufficient magnitude to appear clearly, and the mode-converted wave is both low in magnitude and obscured by the strong Rayleigh wave.

Prior to cross-sectional imaging, the B-scans, as part of the TOFDI process [142, 143, 163], are band-pass filtered, and have horizontal and angled linear features (direct and reflected surface waves respectively) automatically removed from the image [143], the result of which is figure 8.34. Although this process makes the scattered shear wave relatively stronger, the scattered compression and mode-converted waves are unfortunately still not sufficiently visible. For B-scan display purposes, the reliability mapping has been used with the maximum suppression (as discussed in chapter 6), but this was not necessary when performing the cross-sectional imaging.

The cross-sectional image formed from the B-scans, with (figure 8.35a) and without (figure 8.35b) the processing to remove waves that have not been scattered from a defect, are shown with magnitude as the luminance for clarity. The greyscale axis is scaled for each figure such that zero on the scale always represents zero in the data, and one on the scale represents the highest

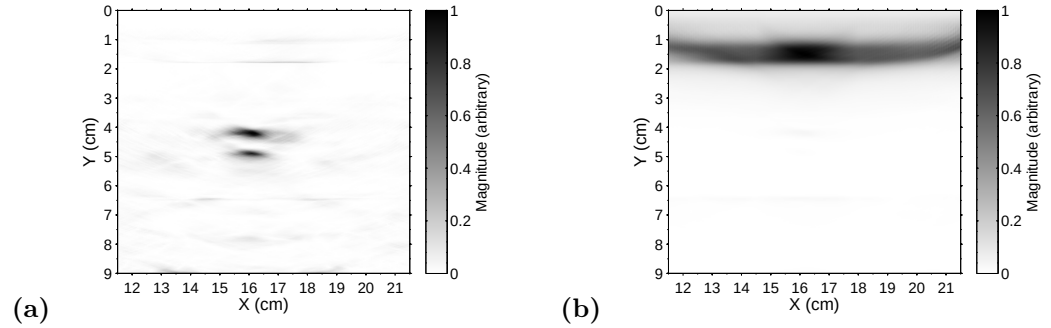


Figure 8.35: TOFDI cross-sectional imaging output using figure 8.34 as input (a) and comparing with using figure 8.33 as input (b). Using the processed B-scan (a), indications representing the top and bottom of the defect are clearly visible, and the defect could be positioned and sized based on this. Using the B-scan without processing (b), there is only an image artefact due to the strongly received surface waves that (in this case, and for TOFDI generally) contains no useful defect information.

magnitude point for that data set. Minimising the non-scattered waves from the B-scan has clearly allowed the processing to form a useful cross-sectional image, in which the top and bottom of the defect can be clearly seen. Without the B-scan processing, all that is seen is an image artefact due to the strong signal from the directly received Rayleigh wave. Noting the earlier statements regarding the top and bottom signals from a side-drilled hole (section 8.2.5), the position of the top of the defect in figure 8.35a is correctly identified by the automatic peak finder, as being at a depth of 42mm (the 45mm centre offset by the 3mm radius). The bottom should be at 48mm, but appears at 49mm; this is not due to a problem with the automatic peak finder, but is due to the extra time taken by the waves to creep round the defect.

8.4.1 Comparing against parabola filtering

The parabola filtering technique (chapter 7) provides similar output to the technique described in this chapter. Consequently, it is worth comparing the quality of the output, in terms of how clear defect detections are, and the speed of the processing, using the same experimental data set already used in this section. To make the comparison simpler to interpret, only a single wave type is considered, the shear wave, and only a single B-scan, but this will mean that there are more artefacts in the output, and the SNR will be lower.

In this case, transducer separation is $s = 7.8$cm, the shear wave speed is $c_s = 3226$m/s, and the pulse period is $T = 1\mu$s (approximately, but this is the value used for T within the parabola filtering algorithm). The sampling frequency is more than sufficient at 10MHz (downsampled from 100MHz after application of a 2MHz low-pass filter). The parabola in the data is entirely contained within the scanning range $-s/2 \leq x \leq s/2$. The depth range chosen for this comparison is 1cm to 8cm. The grid resolution has been chosen as 0.5mm, which is sufficient for good quality cross-sectional imaging; it is one tenth the wavelength, if the maximum frequency were 645.2kHz (this exact number has been chosen for numerical convenience), which is above the frequency at which the majority of the signal energy appears for scattered waves. Horizontal and angled linear features (direct and reflected surface waves respectively) have been automatically removed from the image, but reliability mapping has not been used at all, including in the display of the B-scan, figure 8.36.

The parabola filtered B-scan (figure 8.36a) and the cross-sectional image (figure 8.36b) formed

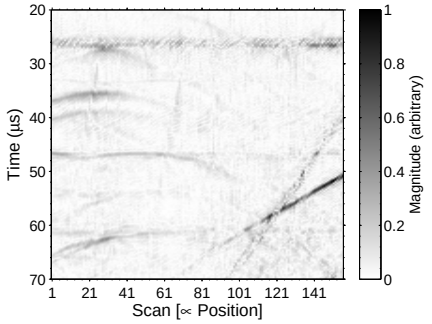


Figure 8.36: Automatically processing figure 8.33 to remove horizontal and angled linear features, converts figure 8.33a to this B-scan. However, unlike with figure 8.34a, reliability mapping has not been used for this display. The input is ready to be processed separately by the parabola filtering stage and the cross-sectional imaging stage.

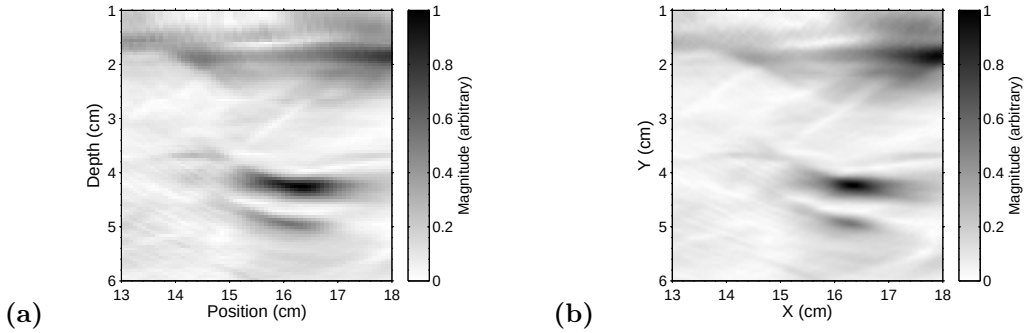


Figure 8.37: Parabola filtering (a), and cross-sectional imaging (b), are compared, with both using figure 8.36 as input after conversion to complex data. The outputs are very similar, and indications representing the top and bottom of the defect are visible in both cases.

from the input B-scan, after conversion to the analytic signal form, are shown with magnitude as the luminance. As before, the greyscale axis is scaled for each figure such that zero on the scale always represents zero in the data, and one on the scale represents the highest magnitude point for that data set. For display purposes, only a small region of the output is shown, including a depth range of 1cm to 6cm, and a positional range (here the horizontal axis) of 13cm to 18cm; there were no significant artefacts or defect indications outside of the range shown. The output is almost identical; the top and bottom of the defect are visible, the SNR is very similar, and the difference between the normalised magnitude for the top and bottom peaks is also very similar.

The processing took place on a desktop machine running MATLAB R2007b (32bit edition) on Windows 7 (64bit edition). The PC had an Intel Core2 CPU, model 6420, at 2.13GHz, 2GB of RAM (more than necessary for this test), and no other CPU intensive applications running concurrently. Averaged over 1000 runs, excluding one-off stages that would not need to be repeated if the technique was used as part of a continuously scanning system, and using complex data, the parabola filtering takes 0.036s, and the cross-sectional imaging takes 0.187s, making the cross-sectional imaging over five times slower than the parabola filtering. However, this is highly dependent on the environment in which the test is running, and if implemented in a different programming language, taking more explicit advantage of multiple core processing, it is not clear how this result would change. This time does not include the linear interpolation stage of the parabola filtering technique, as although it has been used here for the purposes of displaying the parabola

filtering, it is optional. If it is deemed to be a necessary part of the parabola filtering, then it would add approximately 0.012s (averaged over 1000 runs), making the parabola filtering only four times faster.

It would seem from this that the parabola filtering technique has the advantage over the cross-sectional imaging technique. However, it is possible to implement the cross-sectional imaging so that it can deal with issues such as uneven scan intervals and anisotropy, without much additional work. The parabola filtering technique would require far more parabola filters to be applied for the latter, reducing the speed advantage, and there is no clear way in which the former could be implemented. If a sample has inhomogeneity of a form known about, this could easily be incorporated into cross-sectional imaging, but again, it is not clear how parabola filtering could be adapted to this problem. Therefore, although cross-sectional imaging is slower than parabola filtering, it does have additional flexibility that might prove useful when using the technique in an industrial environment.

8.5 Future work

The samples tested have been very limited in number and type, and the sensible next step is to test a greater range of samples, containing a greater range of defects. Simulated defects should include side-drilled holes, bottom-drilled holes, back-wall breaking slots, and surface (front-wall) breaking slots. Real defects should include back-wall breaking cracks, surface (front-wall) breaking cracks, internal cracks, and more volumetric defects such as voids. This should be done for a range of steels, and it would also be beneficial to scan samples that are at high temperature. As the temperature is likely to vary within the bulk, models of temperature distribution would need to be combined with ultrasonic wave speed data to form a velocity map of the sample. Finally, real continuously cast steel billets have cross-sections such as 120mm^2, 140mm^2, 180mm^2, and 210mm^2. The tested samples had a depth of approximately 90mm, and therefore testing a larger cross-section sample would assess if sufficient signal strength remained after scattering off a defect at a depth of, for example, 200mm.

Cross-sectional feature analysis should be expanded and enhanced, to provide automatic defect sizing, and to try and distinguish between defect types, but as already discussed, it is not clear the extent to which different defects could be distinguished, if at all.

Little work has been done towards optimising the array configuration, partly due to the small size of the samples that can be tested safely in the laboratory environment. On larger samples that can be safely tested in industrial environments, greater numbers of EMATs could be deployed.

More speculatively, artefacts could be reduced by using the output of the imaging as an input into a model of the sample. The simulation output could be imaged by TOFDI, and the artefacts observed in the simulated image could be identified and removed from the experimental image. By iterative refinement, the TOFDI image and the model could be improved, until a desired level of artefact reduction was achieved. However, this would most likely be computationally expensive, and difficult to implement.

8.6 Conclusions

TOFDI uses a sparse transducer array (laser-EMAT for non-contact applications), and an imaging stage based on multiple wave mode (and mode-converted) ultrasound scattering by defects. B-

scans are processed to remove direct and reflected surface waves before being passed into the imaging process. The emitter is assumed to approximate a point source, and the receivers are also assumed to approximate points. Scatterers are assumed to be point sources, and a scatterer could lie anywhere on what is sometimes referred to as the travel-time locus curve. TOFDI, like SAFT, locates scatterers by combining all contributions for a point scatterer, for all wave modes and receiver positions.

TOFDI was tested against raycasting simulations, FEM simulations, and experimental data. It has been demonstrated that with only a single emitter and two receivers, a sample can be imaged successfully. The output is a cross-sectional image, and a list of the positions of the defect indications in the image. One of the advantages of the cross-sectional imaging technique is that the output is considerably easier to interpret than the B-scans used as input.

8.7 Use in development of TOFDI

This chapter has presented the method by which the primary outputs of TOFDI, the cross-sectional image and the list of defects, are produced. If a large scan distance is used, the results can be comparable to the TFM technique, if the scans in the latter are not combined to form an overall image. From this, it can be seen that there are applications for which TOFDI would be the most suitable choice, in particular, the prototypical application of testing continuously cast steel products. The speed advantage of the parabola filtering technique is also clear, although this is at the expense of flexibility in the future development of TOFDI.

Chapter 9

Conclusions

This chapter summarises the conclusions made regarding the various parts of TOFDI, stating what has been developed, what each stage does, and several other considerations.

9.1 Summary

TOFDI is based on TOFD, adding a cross-sectional imaging stage by exploiting the scattering of ultrasonic waves from bulk defects in metals. Using a sparse transducer array, a laser ultrasound ablative source, and a sparse array of receiving EMATs for non-contact (linear) scanning, multiple wave modes are emitted and received, with mode-conversions whenever waves are scattered. Standard signal processing techniques reduce noise, such as band-pass filters. A B-scan is formed from multiple data captures (A-scans), with time and scan position axes and colour representing amplitude or magnitude. The prototype application is testing rough steel products of continuous casting, but the techniques created to facilitate operation of TOFDI are applicable elsewhere.

B-scans may contain strong features, due to surface waves propagating directly from emitter to receiver, and after a reflection off a surface edge, which can mask scattered bulk waves. Reflected surface waves can form lines in the B-scan at angles less than 45° from the time axis, which will contain gaps and are considered sparse unless the ultrasound pulse is temporally very broad. The problem is worse for temporally sharper signals or coarser scanning movement, and is still a problem for relatively small scanning increments of 0.8mm and surface waves that at 1µs are relatively long. Such B-scans are produced by the laser-EMAT system used within this project. It has been demonstrated that direct surface waves can be dealt with very simply by subtracting the median along the scan axis at each time point, and clamping the result such that the magnitude of any specific point cannot be larger than its previous value, although the sign of the amplitude can change. A standard HT is not designed to deal with sparse lines, but the modified HT incorporates the constraint that a sparse line can exist along the time dimension by rescaling the B-scan, and it has been demonstrated that this successfully removes reflected surface waves from B-scan images.

A B-scan contains parabolic features characteristic of a point-like scatterer, which can be effectively identified using human vision. The parabola matched filter technique, developed to perform the task of transforming parabolas to peaks, is simple to implement. It relies upon the parabolas having a temporal width such that the shape of the matched filter can change to some extent without moving outside the parabola to be filtered, and a spatial width sufficient for the matched filter to be able to average a significant number of points. The allowance for some change in shape

allows a small number of matched filters to effectively filter a large range of depths, and hence the procedure can be relatively fast. The peaks, that are the response to the matched filter having found a parabola, have a finite size, dominated by the temporal width of the feature. The matched filter also reduces the magnitude of random noise, and to some extent non-parabolic features, that are present in the B-scan. After filtering, the scan can be displayed as depth versus lateral position. Operation of the technique against experimental and simulated data of parabolas formed by shear waves scattering from defects has been demonstrated. The technique worked very well in the case of the simulated data, as might be expected; the parabolas were spatially long, and this meant that the filter has a significant length of parabola over which to operate such that AWGN is reduced significantly. This suggests that detection of spatially long parabolas could be performed using this method, even in the presence of considerable noise. In the experimental case, although the parabolas had the same temporal width as the simulated case, they were spatially very short (not much of the parabola was present in the scan), and this reduced the effectiveness of the parabola matched filter. Despite this, the SNR showed a good improvement, the parabolas were still formed into peaks, and the depth could easily be read from the scale, which is a useful function for users not adept at assessing the depth of a scatterer based on a parabola. Although not demonstrated, the technique also works against parabolas formed through mode-conversion at the scattering point. Care must be taken not to confuse a parabola formed by this process with one formed without mode-conversion, as the response to the indication is shifted towards the side in which the slower wave is present in the mode-converted case, and also appears at an incorrect depth. If the signal is temporally quite sharp or the parabola spatially quite long, this would not be a serious issue, as the difference between the standard and mode-converted parabolas would mean that the indication in the mode-converted parabola matched filter response would be significantly smaller than the response from the standard parabola matched filter. However, for temporally wider signals that are spatially short, the difference can make the correct response ambiguous. Scatterers not exactly in the plane perpendicular to the surface containing the emitter and receiver can still be filtered using this technique. However, the response will be defocused and smaller than for a scatterer in the plane. Scatterers at a considerable perpendicular distance from the plane would not be filtered correctly, but this is mitigated to a large extent by the emitter and receiver used for these tests (and indeed typically) having a main beam that is primarily in the plane. Multiple B-scans are combined to reduce noise further.

The B-scan, without application of the parabola matched filter, is processed to form a cross-sectional image, such that multiple defects can easily be detected, identified, positioned, and potentially sized, including those near the surface and further within the bulk. The emitter is assumed to approximate a point source, and the receivers are also assumed to approximate points. Scatterers are assumed to be point sources, and a scatterer could lie anywhere on what is sometimes referred to as the travel-time locus curve. TOFDI, like SAFT, locates scatterers by combining all contributions for a point scatterer, for all wave modes and receiver positions. TOFDI was tested against raycasting simulations, FEM simulations, and experimental data. It has been demonstrated that with only a single emitter and two receivers, a sample can be imaged successfully. The output is a cross-sectional image, and a list of the positions of the defect indications in the image. One of the advantages of the cross-sectional imaging technique is that the output is considerably easier to interpret than the B-scans used as input.

Any transducer has a finite size for both generation and detection, and the affect of non-zero size on frequency response must be considered. Models were constructed for both far-field and

near-field cases. Simulated data was found to agree with experimental tests when considering nulls in the frequency response. The models were too narrow in scope to fully describe the impact of changing coil configuration on the frequency response; for example, they do not include frequency changes due to changes in mutual and self inductance.

The standard phase correlation technique has been applied to track the relative position between transducer and sample surface. The phase correlation technique is simple to implement, using the FFT for computational speed, and is able to detect translation in two dimensions and (optionally) rotation, despite the presence of significant noise. It has been extended to sub-pixel precision by conventional FFT up-sampling, and filtering of spatial frequencies outside of those representing real image components reduces the impact of AWGN. The algorithm has been tested using only a basic low resolution camera (640×480, fixed focus), and linear sample movement has been determined to sub-pixel precision on experimental data, confirmed manually as a median accuracy of 0.01mm of linear movement (0.06 of a pixel) in this case, despite uneven illumination. The performance as noise increased was assessed, and for simulated noise with a standard deviation a fifth of the mean signal level, injected into the experimental data, the median accuracy was 0.02mm of linear movement (0.13 of a pixel) in this case, with some variation in the results due to the random nature of the noise introduced. Rotation was determined to 0.1° on simulated data, but the limits of resolution and noise tolerance were not explored to the same extent as for translational motion.

9.2 Future development of TOFDI

This chapter has summarised the work on TOFDI, but there is still more work required before TOFDI can be be used in an industrial environment.

Starting with data collection, a laser-EMAT system has been shown, in other work [5], to operate in environments with a hot, rough, and moving sample, including in single shot mode (without averaging multiple data captures at the same position). No problems are expected with the rate of data collection, but further tests are required to check that the data is at a sufficient SNR for TOFDI, and how powerful a laser is required. In addition, it is necessary to check how much cooling of the EMAT is required for continual operation, over days, without incurring a substantial reduction in performance, and if build-up of ferromagnetic particles, due to the strong permanent magnet, is a concern. If it is found to be a concern, a removal mechanism may have to be designed, or alternatively, an electromagnet (pulsed if necessary [110]) used instead of a permanent magnet. An electromagnet might allow the particles to naturally drop away in the short time between each sample being scanned, and might also prove more robust to the high temperatures experienced, if appropriately cooled. Various array configurations, and improvements in EMAT design to take advantage of them, should be developed. Filtering before the data is digitised can improve the dynamic range of the data, as less of the range needs to be surrendered to noise, as well as improving the SNR, and the best filters to achieve this should be explored.

The data processing, including the cross-sectional imaging, is relatively fast, but it could be made considerably faster. There are several stages that are currently just for debugging, or have not been optimised for continuous running, that could be changed in a system used in industry. In addition, many modern computer systems have multiple CPU cores, and as several parts of TOFDI are suitable for parallelisation, significant speed gains could be made. Not all stages of TOFDI have to run to produce a cross-sectional image, and some can be removed without negatively impacting the final output, depending on the conditions under which data is collected. For example, it is not

necessary to remove reflected surface waves, if such waves are not observed in the B-scan, perhaps because the sample is not scanned close to the edges. Alternatively, if such features always appear in a similar place in the B-scan, considerable optimisations can be made. Parameters can be optimised to improve speed and the quality of the outputs.

The data output and visualisation is simple enough to interpret, and a list of defects with positions is also output. This is fine for a small number of samples, but if a large number of samples are being continuously scanned, a more sophisticated system for data visualisation may be required. As an example, the list of defects could be logged, and prioritised depending on the position within the sample, and how defective this makes the sample. A system to alert a user regarding problem cases could be developed. Upon selecting a sample to examine more closely, then the cross-sectional image could be displayed, with the defect positions clearly indicated.

TOFDI is primarily for bulk defects, but surface breaking cracks can also be assessed, and additional information may be gleaned by examining the surface waves, as has been done when checking rail [61–64, 105].

These steps towards use in an industrial environment will require close collaboration with an industrial partner, and cannot be achieved purely in a laboratory environment.

References

[1] J. P. Charlesworth and J. A. G. Temple, *Engineering Applications of Ultrasonic Time-of-Flight Diffraction*, 2nd ed., ser. Ultrasonic Inspection In Engineering, M. J. Whittle, Ed. Baldock, UK: Research Studies Press, 2001.

[2] J. A. G. Temple, "Time-of-flight inspection: theory," *Nuclear Energy*, vol. 22, no. 5, pp. 335–348, Oct. 1983.

[3] J. Blitz and G. Simpson, *Ultrasonic Methods of Non-destructive Testing*, ser. Non-destructive Evaluation. London, UK: Chapman and Hall, 1996.

[4] X. Jian, I. Baillie, and S. Dixon, "Steel billet inspection using laser-EMAT system," *Journal of Physics D: Applied Physics*, vol. 40, no. 5, pp. 1501–1506, Mar. 2007.

[5] I. Baillie, P. Griffith, X. Jian, and S. Dixon, "Implementing an ultrasonic inspection system to find surface and internal defects in hot, moving steel using EMATs," *Insight*, vol. 49, no. 2, pp. 87–92, Feb. 2007.

[6] C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications.* Bristol, UK: Adam Hilger, 1990.

[7] C. B. Scruby, "Some applications of laser ultrasound," *Ultrasonics*, vol. 27, no. 4, pp. 195–209, July 1989.

[8] N. F. Haines, "Progress towards an understanding of reliability in non-destructive evaluation," *Nuclear Energy*, vol. 22, no. 5, pp. 349–356, Oct. 1983.

[9] D. E. Bray and R. K. Stanley, *Nondestructive Evaluation: A Tool in Design, Manufacturing, and Service.* Boca Raton, USA: CRC Press, 1997.

[10] R. A. Murgatroyd, P. J. Highmore, S. F. Burch, T. Bann, and A. T. Ramsey, "PISC-II parametric study on flaw characterization using the tandem and TOFD techniques," *International Journal of Pressure Vessels and Piping*, vol. 35, no. 1-4, pp. 137–169, 1988.

[11] M. G. Silk and B. H. Lidington, "Defect sizing using an ultrasonic time delay approach," *British Journal of Non-Destructive Testing*, vol. 17, no. 2, pp. 33–36, Mar. 1975.

[12] M. G. Silk and B. H. Lidington, "The potential of scattered or diffracted ultrasound in the determination of crack depth," *Non-destructive Testing*, vol. 8, no. 3, pp. 146–151, June 1975.

[13] B. H. Lidington, M. G. Silk, P. Montgomery, and G. Hammond, "Ultrasonic measurements of the depth of fatigue cracks," *British Journal of NDT*, vol. 18, pp. 165–170, Nov. 1976.

[14] L. N. J. Poulter, "Signal processing methods applied in the ultrasonic inspection of PWR inlet nozzles," *NDT International*, vol. 19, no. 3, pp. 141–144, June 1986.

[15] B. H. Lidington, D. H. Saunderson, and M. G. Silk, "Interference effects in the reflection of ultrasound from shallow slits," *Non-destructive Testing*, vol. 8, no. 4, pp. 185–190, Aug. 1975.

[16] I. Atkinson, R. J. Hudgello, and L. N. J. Poulter, "High-accuracy defect sizing by the use of in-situ monitoring techniques," in *Proceedings of the 12th International Conference on NDE in the Nuclear and Pressure Vessel Industries*, Philadelphia (Pennsylvania), USA, July 1994, pp. 287–296.

[17] P. Bolland, L. F. C. L. Y. Voon, B. Gremillet, L. Pillet, A. Diou, and P. Gorria, "The application of Hough transform on ultrasonic images for the detection and characterization of defects in non-destructive inspection," in *Proceedings of the 3rd International Conference on Signal Processing (ICSP '96)*, Oct. 1996, pp. 393–396.

[18] K. Maalmi, R. Benslimane, L. F. C. L. Y. Voon, and E. Fauvet, "Towards automatic analysis of ultrasonic time-of-flight diffraction data using genetic-based inverse Hough transform," *Insight*, vol. 51, no. 4, pp. 184–191, Apr. 2009.

[19] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_4/2_4_1.html`.

[20] B. M. Hawker, "The automation of inspection of thick steel sections using the ultrasonic time-of-flight technique," *Nuclear Energy - Journal of the British Nuclear Energy Society*, vol. 22, no. 5, pp. 309–318, 1983.

[21] A. Rogerson, P. J. Highmore, and L. N. J. Poulter, "Defect sizing using automated ultrasonic inspection techniques at Risley-Nuclear-Laboratories," *Nuclear Energy - Journal of the British Nuclear Energy Society*, vol. 22, no. 5, pp. 367–378, 1983.

[22] P. J. Highmore, A. Rogerson, and L. N. J. Poulter, "The ultrasonic inspection of PISC-II plate-2 by the Risley-Nuclear-Laboratories," *British Journal of Non-Destructive Testing*, vol. 30, no. 1, pp. 9–17, Jan. 1988.

[23] A. Rogerson, L. N. J. Poulter, P. Clough, and A. G. Cooper, "RNL automated ultrasonic inspection of the PISC-II PWR inlet nozzle (plate-3)," *British Journal of Non-Destructive Testing*, vol. 30, no. 2, pp. 86–93, Mar. 1988.

[24] S. K. Nath, K. Balasubramaniam, C. V. Krishnamurthy, and B. H. Narayana, "Sizing of surface-breaking cracks in complex geometry components by ultrasonic time-of-flight diffraction (ToFD) technique," *Insight*, vol. 49, no. 4, pp. 200–206, Apr. 2007.

[25] M. G. Silk, "The interpretation of TOFD data in the light of ASME-XI and similar rules," *British Journal of Non-Destructive Testing*, vol. 31, no. 5, pp. 242–251, May 1989.

[26] M. W. Stringfellow and J. K. Perring, "Detection and sizing of inner radius defects in DDT plate 4 (simulated PWR nozzle) by the ultrasonic time-of-flight diffraction technique," *British Journal of NDT*, vol. 26, no. 2, pp. 84–91, Feb. 1984.

[27] M. G. Silk, "Estimates of the probability of detection of flaws in TOFD data with varying levels of noise," *Insight*, vol. 38, no. 1, pp. 31–36, Jan. 1996.

[28] M. G. Silk, "An evaluation of the performance of the TOFD technique as a means of sizing flaws, with particular reference to flaws with curved profiles," *Insight*, vol. 38, no. 4, pp. 280–287, Apr. 1996.

[29] I. Atkinson, "High-accuracy defect sizing for CRDM penetration adapters using the ultrasonic TOFD technique," *Insight*, vol. 37, no. 3, pp. 175–178, Mar. 1995.

[30] L. N. J. Poulter, A. Rogerson, A. J. Willetts, and A. V. Dyke, "Inspection of the defect detection trials plate 4 by the materials physics department, RNL," *British Journal of NDT*, vol. 26, no. 2, pp. 77–83, Feb. 1984.

[31] J. P. Charlesworth and B. M. Hawker, "Inspection of the near-surface defect plate (DDT3) by the ultrasonic time-of-flight technique," *British Journal of NDT*, vol. 26, no. 2, pp. 106–112, Feb. 1984.

[32] M. G. Silk, "Benefits of signal-processing in ultrasonic inspection," *Insight*, vol. 36, no. 10, pp. 776–781, Oct. 1994.

[33] L. Capineri, P. Grande, L. Masotti, C. G. Windsor, and J. A. G. Temple, "Advanced image processing techniques for automatic interpretation of time-of-flight diffraction images," in *Acoustical Imaging: Proceedings of the 23rd International Symposium on Acoustical Imaging*, ser. Acoustical Imaging, S. Lees and L. A. Ferrari, Eds., vol. 23, The Acoustical Society of America. Boston (Massachusetts), USA: Plenum Press, Apr. 1997, pp. 421–425.

[34] K. Maalmi, A. E. Ouaazizi, R. Benslimane, L. F. C. L. Y. Voon, A. Diou, and P. Gorria, "Detecting parabolas in ultrasound B-scan images with genetic-based inverse voting Hough transform," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002 (ICASSP '02)*, vol. 4, 2002, pp. 3337–3340.

[35] A. Al-Ataby, W. Al-Nuaimy, and O. Zahran, "Towards automatic flaw sizing using ultrasonic time-of-flight diffraction," *Insight*, vol. 52, no. 7, pp. 366–371, July 2010.

[36] S. F. Burch and A. T. Ramsey, "Digital-techniques to derive and display the phases of ultrasonic pulses," *British Journal of Non-Destructive Testing*, vol. 31, no. 4, pp. 185–188, Apr. 1989.

[37] S. F. Burch, A. R. Lomas, and A. T. Ramsey, "Practical automated ultrasonic characterization of welding defects," *British Journal of Non-Destructive Testing*, vol. 32, no. 7, pp. 347–350, July 1990.

[38] L. Capineri, H. G. Tattersall, J. A. G. Temple, and M. G. Silk, "Time-of-flight diffraction tomography for NDT applications," *Ultrasonics*, vol. 30, no. 5, pp. 275–288, Feb. 1992.

[39] S. F. Burch, "Comparison of SAFT and two-dimensional deconvolution methods for the improvement of resolution in ultrasonic B-scan images," *Ultrasonics*, vol. 25, no. 5, pp. 259–266, Sep. 1987.

[40] C. Dazhao, G. Tie, Y. Yingxue, and Y. Yuan, "Weld crack testing method using ultrasonic TOFD technique," *Insight*, vol. 52, no. 4, pp. 188–191, Apr. 2010.

[41] K. Nagai, "A new synthetic-aperture focusing method for ultrasonic B-scan imaging by the Fourier transform," *IEEE Transactions on Sonics and Ultrasonics*, vol. 32, no. 4, pp. 531–536, July 1985.

[42] K. J. Langenberg, M. Berger, T. Kreutter, K. Mayer, and V. Schmitz, "Synthetic aperture focusing technique signal processing," *NDT International*, vol. 19, no. 3, pp. 177–189, June 1986.

[43] S. Boonsang, J. Zainal, and R. J. Dewhurst, "Synthetic aperture focusing techniques in time and frequency domains for photoacoustic imaging," *Insight*, vol. 46, no. 4, pp. 196–199, Apr. 2004.

[44] F. Lingvall, T. Olofsson, and T. Stepinski, "Synthetic aperture imaging using sources with finite aperture: Deconvolution of the spatial impulse response," *The Journal of the Acoustical Society of America*, vol. 114, no. 1, pp. 225–234, July 2003.

[45] S. D. Silverstein and L. J. Thomas, "Analytical comparison of sensor signal processing enhancements for NDT synthetic aperture ultrasonic imaging," *IEEE Transactions on Image Processing*, vol. 2, no. 1, pp. 60–67, Jan. 1993.

[46] P. C. Snowdon, S. Johnstone, and S. Dewey, "A 2D static ultrasonic array of passive probes for improved probability of detection," *Nondestructive Testing and Evaluation*, vol. 19, no. 3, pp. 111–120, Sep. 2003.

[47] P. C. Snowdon, S. Johnstone, and S. Dewey, "Improving the probability of detection for a 2D ultrasonic array using digital signal processing techniques," *Insight*, vol. 45, no. 11, pp. 743–745, Nov. 2003.

[48] G. Baskaran, K. Balasubramaniam, and C. L. Rao, "Shear-wave time of flight diffraction (S-TOFD) technique," *NDT & E International*, vol. 39, no. 6, pp. 458–467, Sep. 2006.

[49] S. B. Palmer and S. Dixon, "Industrially viable non-contact ultrasound," *Insight*, vol. 45, no. 3, pp. 211–217, Mar. 2003.

[50] P. R. Murray and R. J. Dewhurst, "A laser/EMAT system for thickness monitoring applications using shear and L-S mode-converted waves," *Measurement Science and Technology*, vol. 12, no. 10, pp. 1651–1659, Oct. 2001.

[51] S. Dixon, C. Edwards, and S. B. Palmer, "A laser-EMAT system for ultrasonic weld inspection," *Ultrasonics*, vol. 37, no. 4, pp. 273–281, Apr. 1999.

[52] R. J. Dewhurst, C. Edwards, and S. B. Palmer, "Noncontact detection of surface-breaking cracks using a laser acoustic source and an electromagnetic acoustic receiver," *Applied Physics Letters*, vol. 49, no. 7, pp. 374–376, Aug. 1986.

[53] J. R. Bernstein and J. B. Spicer, "Line source representation for laser-generated ultrasound in aluminum," *Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1352–1357, Mar. 2000.

[54] A. Idris, C. Edwards, and S. B. Palmer, "Acoustic wave measurements at elevated temperature using a pulsed laser generator and an electromagnetic acoustic transducer detector," *Nondestructive Testing and Evaluation*, vol. 11, no. 4, pp. 195–213, 1994.

[55] Y. Fan, S. Dixon, R. S. Edwards, and X. Jian, "Ultrasonic surface wave propagation and interaction with surface defects on rail track head," *NDT & E International*, vol. 40, no. 6, pp. 471–477, Sep. 2007.

[56] X. Jian, S. Dixon, K. T. V. Grattan, and R. S. Edwards, "A model for pulsed Rayleigh wave and optimal EMAT design," *Sensors and Actuators A: Physical*, vol. 128, no. 2, pp. 296–304, Apr. 2006.

[57] J. L. Blackshire and S. Sathish, "Near-field ultrasonic scattering from surface-breaking cracks," *Applied Physics Letters*, vol. 80, no. 18, pp. 3442–3444, May 2002.

[58] X. Jian, Y. Fan, R. S. Edwards, and S. Dixon, "Surface-breaking crack gauging with the use of laser-generated Rayleigh waves," *Journal of Applied Physics*, vol. 100, no. 6, p. 064907, Sep. 2006.

[59] R. J. Dewhurst, C. Edwards, A. D. W. McKie, and S. B. Palmer, "Estimation of the thickness of thin metal sheet using laser generated ultrasound," *Applied Physics Letters*, vol. 51, no. 14, pp. 1066–1068, Oct. 1987.

[60] S. Dixon, C. Edwards, and S. B. Palmer, "High accuracy non-contact ultrasonic thickness gauging of aluminium sheet using electromagnetic acoustic transducers," *Ultrasonics*, vol. 39, no. 6, pp. 445–453, Oct. 2001.

[61] S. Dixon, R. S. Edwards, and X. Jian, "Inspection of rail track head surfaces using electromagnetic acoustic transducers (EMATs)," *Insight*, vol. 46, no. 6, pp. 326–330, June 2004.

[62] R. S. Edwards, S. Dixon, and X. Jian, "Enhancement of the Rayleigh wave signal at surface defects," *Journal of Physics D: Applied Physics*, vol. 37, no. 16, pp. 2291–2297, Aug. 2004.

[63] R. S. Edwards, S. Dixon, and X. Jian, "Depth gauging of defects using low frequency wideband Rayleigh waves," *Ultrasonics*, vol. 44, no. 1, pp. 93–98, Jan. 2006.

[64] R. S. Edwards, S. Dixon, and X. Jian, "Characterisation of defects in the railhead using ultrasonic surface waves," *NDT & E International*, vol. 39, no. 6, pp. 468–475, Sep. 2006.

[65] X. Jian, S. Dixon, I. Baillie, R. Edwards, and J. Morrison, "Integrity evaluation of steel products using EMATs," *Journal of Physics D: Applied Physics*, vol. 40, no. 2, pp. 300–304, Jan. 2007.

[66] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, "Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation," *NDT & E International*, vol. 38, no. 8, pp. 701–711, Dec. 2005.

[67] S. F. Burch, N. J. Collett, S. Terpstra, and M. V. Hoekstra, "M-skip: a quantitative technique for the measurement of wall loss in inaccessible components," *Insight*, vol. 49, no. 4, pp. 190–194, Apr. 2007.

[68] T. Hutt and F. Simonetti, "Reconstructing the shape of an object from its mirror image," *Journal of Applied Physics*, vol. 108, no. 6, p. 064909, Sep. 2010.

[69] J. P. Charlesworth and J. A. G. Temple, "Ultrasonic time-of-flight inspection through anisotropic cladding," in *Proceedings of the Conference on Periodic Inspection of Pressurised Components*, ser. Institute of Mechanical Engineers Conference Publications, London, UK, Oct. 1982, pp. 117–124.

[70] C. Birac, S. Crutzen, M. Certo, P. Dombret, R. A. Murgatroyd, G. Deuster, P. J. Highmore, S. F. Burch, T. Bann, A. T. Ramsey, T. Cola, G. Mariani, F. Lakestani, E. Borloo, and F. Merli, "The PISC-II parametric studies program - discussion on session-II," *International Journal of Pressure Vessels and Piping*, vol. 35, no. 1-4, pp. 247–258, 1988.

[71] K. F. Graff, *Wave Motion in Elastic Solids*. New York, USA: Dover Publications, 1975.

[72] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_2/2_2_2.html`.

[73] I. Arias and J. D. Achenbach, "Thermoelastic generation of ultrasound by line-focused laser irradiation," *International Journal of Solids and Structures*, vol. 40, no. 25, pp. 6917–6935, Dec. 2003.

[74] M. G. Silk, *Ultrasonic Transducers for Nondestructive Testing*. Bristol, UK: Adam Hilger Ltd., 1984.

[75] J. Krautkramer and H. Krautkramer, *Ultrasonic Testing of Materials*, 4th ed. Berlin, Germany: Springer-Verlag, 1990.

[76] H. M. Ledbetter and J. C. Moulder, "Laser-induced Rayleigh waves in aluminum," *Journal of the Acoustical Society of America*, vol. 65, no. 3, pp. 840–842, Mar. 1979.

[77] J. D. Achenbach, *Wave Propagation in Elastic Solids*, ser. North-Holland series in applied mathematics and mechanics. London, UK: North-Holland, 1973.

[78] B. A. Auld, *Acoustic fields and waves in solids*, 2nd ed. Florida, USA: Krieger Publishing Company, 1990, ch. Reflection and refraction, pp. 30–38.

[79] L. Rayleigh, "On waves propagated along the plane surface of an elastic solid," *Proceedings of the London Mathematical Society*, vol. 17, no. 1, pp. 4–11, Nov. 1885.

[80] S. Dixon and S. B. Palmer, "Wideband low frequency generation and detection of Lamb and Rayleigh waves using electromagnetic acoustic transducers (EMATs)," *Ultrasonics*, vol. 42, no. 10, pp. 1129–1136, Aug. 2004.

[81] R. G. Steg and P. G. Klemens, "Scattering of Rayleigh waves by surface irregularities," *Physical Review Letters*, vol. 24, no. 8, pp. 381–383, Feb. 1970.

[82] A. A. Maradudin and D. L. Mills, "Attenuation of Rayleigh surface waves by surface roughness," *Applied Physics Letters*, vol. 28, no. 10, pp. 573–575, May 1976.

[83] A. A. Maradudin and D. L. Mills, "The attenuation of Rayleigh surface waves by surface roughness," *Annals of Physics*, vol. 100, no. 1-2, pp. 262–309, 1976.

[84] R. F. Wallis, D. L. Mills, and A. A. Maradudin, "Attenuation of Rayleigh waves by point defects," *Physical Review B*, vol. 19, no. 8, pp. 3981–3995, Apr. 1979.

[85] R. V. Goldstein and J. Lewandowski, "Surface roughness induced attenuation and changes in the propagation velocity of long Rayleigh-type waves," *Acta Mechanica*, vol. 91, no. 3-4, pp. 235–243, 1992.

[86] C. Pecorari, "Attenuation and dispersion of Rayleigh waves propagating on a cracked surface: an effective field approach," *Ultrasonics*, vol. 38, no. 1-8, pp. 754–760, Mar. 2000.

[87] E. Hu, Y. He, and Y. Chen, "Experimental study on the surface stress measurement with Rayleigh wave detection technique," *Applied Acoustics*, vol. 70, no. 2, pp. 356–360, Feb. 2009.

[88] E. P. Papadakis, "Revised grain-scattering formulas and tables," *The Journal of the Acoustical Society of America*, vol. 37, no. 4, pp. 703–710, Apr. 1965.

[89] M. Hirao and H. Ogi, *EMATS for Science and Industry: Noncontacting Ultrasonic Measurements*. London, UK: Kluwer Academic Publishers, 2003.

[90] A. H. Harker, P. Schofield, B. P. Stimpson, R. G. Taylor, and J. A. G. Temple, "Ultrasonic propagation in slurries," *Ultrasonics*, vol. 29, no. 6, pp. 427–438, Nov. 1991.

[91] R. L. Roderick and R. Truell, "The measurement of ultrasonic attenuation in solids by the pulse technique and some results in steel," *Journal of Applied Physics*, vol. 23, no. 2, pp. 267–279, Feb. 1952.

[92] M. Karaoguz, N. Bilgutay, and B. Onaral, "Modeling of scattering dominated ultrasonic attenuation using power-law function," in *IEEE Ultrasonics Symposium Proceedings*, ser. Ultrasonics, vol. 1-2, 2000, pp. 793–796.

[93] E. P. Papadakis, "Ultrasonic attenuation caused by scattering in polycrystalline metals," *The Journal of the Acoustical Society of America*, vol. 37, no. 4, pp. 711–717, Apr. 1965.

[94] J. L. Taylor, *Basic Metallurgy for Non-Destructive Testing*, 2nd ed. Northampton, UK: The British Institute of Non-Destructive Testing, 1996.

[95] E. P. Papadakis, "Ultrasonic attenuation and velocity in three transformation products in steel," *Journal of Applied Physics*, vol. 35, no. 5, pp. 1474–1482, May 1964.

[96] A. A. Vakulenko and V. Y. Kreinovich, "Physico-geometric investigation of brittle fracture during creep," *Journal of Applied Mathematics and Mechanics*, vol. 53, no. 5, pp. 660–665, 1989.

[97] J. M. Coffey and R. K. Chapman, "Application of elastic scattering theory for smooth flat cracks to the quantitative prediction of ultrasonic defect detection and sizing," *Nuclear Energy*, vol. 22, no. 5, pp. 319–333, Oct. 1983.

[98] J. K. Naa and J. L. Blackshire, "Interaction of Rayleigh surface waves with a tightly closed fatigue crack," *NDT & E International*, vol. 43, no. 5, pp. 432–439, July 2010.

[99] S. Dixon, T. J. Harrison, and P. A. Petcher, "Phase changes of ultrasonic bulk waves through focusing measured using a noncontact ultrasonic method," *Applied Physics Letters*, vol. 97, no. 5, p. 054101, Aug. 2010.

[100] S. Dixon, C. Edwards, and S. B. Palmer, "The analysis of adhesive bonds using electromagnetic acoustic transducers," *Ultrasonics*, vol. 32, no. 6, pp. 425–430, Nov. 1994.

[101] S. Dixon, C. Edwards, and S. B. Palmer, "A technique for accurate shear wave velocity measurements of thin epoxy resin samples using electromagnetic acoustic transducers (EMATs)," *Measurement Science and Technology*, vol. 12, no. 5, pp. 615–621, May 2001.

[102] S. Dixon, C. Edwards, and S. B. Palmer, "Texture measurements of metal sheets using wideband electromagnetic acoustic transducers," *Journal of Physics D: Applied Physics*, vol. 35, no. 8, pp. 816–824, Apr. 2002.

[103] S. Dixon, C. Edwards, and S. B. Palmer, "Recent developments in the characterisation of aluminium sheets using electromagnetic acoustic transducers (EMATs)," *Insight*, vol. 44, no. 5, pp. 274–278, May 2002.

[104] S. Dixon, D. Jaques, and S. B. Palmer, "The development of shear and compression elastic moduli in curing epoxy adhesives measured using non-contact ultrasonic transducers," *Journal of Physics D: Applied Physics*, vol. 36, no. 6, pp. 753–759, Mar. 2003.

[105] R. S. Edwards, X. Jian, Y. Fan, and S. Dixon, "Signal enhancement of the in-plane and out-of-plane Rayleigh wave components," *Applied Physics Letters*, vol. 87, no. 19, p. 194104, Nov. 2005.

[106] S. Huang, W. Zhao, Y. Zhang, and S. Wang, "Study on the lift-off effect of EMAT," *Sensors and Actuators A: Physical*, vol. 153, no. 2, pp. 218–221, Aug. 2009.

[107] M. Hirao and H. Ogi, "Electromagnetic acoustic resonance and materials characterization," *Ultrasonics*, vol. 35, no. 6, pp. 413–421, Sep. 1997.

[108] H. Ogi, E. Goda, and M. Hirao, "Increase of efficiency of magnetostriction SH-wave electromagnetic acoustic transducer by angled bias field: piezomagnetic theory and measurement," *Japanese Journal of Applied Physics Part 1 - Regular Papers Short Notes & Review Papers*, vol. 42, no. 5, pp. 3020–3024, May 2003.

[109] B. Dutton, S. Boonsang, and R. J. Dewhurst, "A new magnetic configuration for a small in-plane electromagnetic acoustic transducer applied to laser-ultrasound measurements: modelling and validation," *Sensors and Actuators A: Physical*, vol. 125, no. 2, pp. 249–259, Jan. 2006.

[110] F. Hernandez-Valle and S. Dixon, "Initial tests for designing a high temperature EMAT with pulsed electromagnet," *NDT & E International*, vol. 43, no. 2, pp. 171–175, Feb. 2010.

[111] X. Jian, S. Dixon, R. S. Edwards, and J. Morrison, "Coupling mechanism of an EMAT," *Ultrasonics*, vol. 44, no. 6, pp. 653–656, June 2006.

[112] K. Kawashima, "Theory and numerical calculation of the acoustic field produced in metal by an electromagnetic ultrasonic transducer," *Journal of the Acoustical Society of America*, vol. 60, no. 5, pp. 1089–1099, Nov. 1976.

[113] R. B. Thompson, *Ultrasonic Measurement Methods*, ser. Physical acoustics. London, UK: Academic Press, 1990, vol. XIX, ch. Physical principles of measurements with EMAT transducers, pp. 157–200.

[114] T. J. Moran and R. M. Panos, "Electromagnetic generation of electronically steered ultrasonic bulk waves," *Journal of Applied Physics*, vol. 47, no. 5, pp. 2225–2227, May 1976.

[115] H. A. Wheeler, "Formulas for the skin effect," *Proceedings of the IRE*, vol. 30, no. 9, pp. 412–424, Sept. 1942.

[116] A. V. Vorst, A. Rosen, and Y. Kotsuka, *RF/Microwave Interaction with Biological Tissues*. New Jersey, USA: John Wiley & Sons, 2006.

[117] E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd ed. New Jersey, USA: Prentice-Hall, 1968.

[118] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_6/2_6_1.html`.

[119] H. Ogi, "Field dependence of coupling efficiency between electromagnetic field and ultrasonic bulk waves," *Journal of Applied Physics*, vol. 82, no. 8, pp. 3940–3949, Oct. 1997.

[120] H. L. Grubin, "Direct electromagnetic generation of compressional waves in metals in static magnetic fields," *IEEE Transactions on Sonics and Ultrasonics*, vol. 17, no. 4, pp. 227–229, Oct. 1970.

[121] K. Kawashima, O. B. Wright, and T. Hyoguchi, "High frequency resonant electromagnetic generation and detection of ultrasonic waves," *Japanese Journal of Applied Physics Part 1 - Regular Papers Short Notes & Review Papers*, vol. 33, no. 5, pp. 2837–2845, May 1994.

[122] R. Ribichini, F. Cegla, P. B. Nagy, and P. Cawley, "Quantitative modeling of the transduction of electromagnetic acoustic transducers operating on ferromagnetic media," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 12, pp. 2808–2817, Dec. 2010.

[123] B. D. Cullity, "Fundamentals of magnetostriction," *Journal of Metals*, vol. 23, no. 1, pp. 35–41, Jan. 1971.

[124] D. Jiles, *Introduction to Magnetism and Magnetic Materials*. London, UK: Chapman and Hall, 1991.

[125] R. B. Thompson, "A model for the electromagnetic generation of ultrasonic guided waves in ferromagnetic metal polycrystals," *IEEE Transactions on Sonics and Ultrasonics*, vol. 25, no. 1, pp. 7–15, Jan. 1978.

[126] R. B. Thompson, "Mechanisms of electromagnetic generation and detection of ultrasonic Lamb waves in iron-nickel alloy polycrystals," *Journal of Applied Physics*, vol. 48, no. 12, pp. 4942–4950, Dec. 1977.

[127] R. Ribichini, F. Cegla, P. B. Nagy, and P. Cawley, "Evaluation of electromagnetic acoustic transducer performance on steel materials," in *Review of Progress in Quantitative Nondestructive Evaluation: Proceedings of the 37th Annual Review of Progress in Quantitative Nondestructive Evaluation*, ser. AIP Conference Proceedings, vol. 1335. San Diego (California), USA: AIP, 2011, pp. 785–792.

[128] C. Kittel, *Introduction to Solid State Physics*, 8th ed. New Jersey, USA: John Wiley & Sons, 2005.

[129] R. Jafari-Shapoorabadi and A. Konrad, "The governing electrodynamic equations of electromagnetic acoustic transducers," *Journal of Applied Physics*, vol. 97, no. 10, p. 10E102, May 2005.

[130] X. Jian, S. Dixon, and R. S. Edwards, "Optimising ultrasonic wideband Rayleigh wave generation by pulsed electromagnetic coils," *Nondestructive Testing and Evaluation*, vol. 20, no. 1, pp. 43–62, Mar. 2005.

[131] X. Jian, S. Dixon, K. Quirk, and K. T. V. Grattan, "Electromagnetic acoustic transducers for in- and out-of plane ultrasonic wave detection," *Sensors and Actuators A: Physical*, vol. 148, no. 1, pp. 51–56, Jan. 2008.

[132] J. F. Ready, *Industrial Applications of Lasers*, 2nd ed. London, UK: Academic Press, 1997.

[133] S. Dixon, C. Edwards, and S. B. Palmer, "Generation of ultrasound by an expanding plasma," *Journal of Physics D: Applied Physics*, vol. 29, no. 12, pp. 3039–3044, Dec. 1996.

[134] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_3/2_3_7.html`.

[135] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_3/2_3_6.html`.

[136] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/3_1/3_1_2.html`.

[137] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_2/2_2_1.html`.

[138] G. W. C. Kaye, "Tables of physical and chemical constants (web edition)," 2004, `http://www.kayelaby.npl.co.uk/general_physics/2_3/2_3_5.html`.

[139] P. Krehl, F. Schwirzke, and A. W. Cooper, "Correlation of stress-wave profiles and the dynamics of the plasma produced by laser irradiation of plane solid targets," *Journal of Applied Physics*, vol. 46, no. 10, pp. 4400–4406, Oct. 1975.

[140] D. W. Gregg and S. J. Thomas, "Momentum transfer produced by focused laser giant pulses," *Journal of Applied Physics*, vol. 37, no. 7, pp. 2787–2789, June 1966.

[141] B. Dutton, S. Boonsang, and R. J. Dewhurst, "Modelling of magnetic fields to enhance the performance of an in-plane EMAT for laser-generated ultrasound," *Ultrasonics*, vol. 44, no. 1, pp. 657–665, Dec. 2006.

[142] P. A. Petcher and S. Dixon, "Use of a sparse array for time of flight diffraction and imaging (ToFDI)," in *Proceedings of the National Seminar & Exhibition on Non-Destructive Evaluation*. Tiruchirapalli, India: NDT.net, 2010, pp. 309–314.

[143] P. A. Petcher and S. Dixon, "Time of flight diffraction and imaging (TOFDI) combining B-scans and cross-sectional imaging," in *Review of Progress in Quantitative Nondestructive Evaluation: Proceedings of the 37th Annual Review of Progress in Quantitative Nondestructive Evaluation*, ser. AIP Conference Proceedings, vol. 1335. San Diego (California), USA: AIP, 2011, pp. 744–751.

[144] S. L. Marple, "Computing the discrete-time analytic signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[145] J. B. Hagan, *Radio-Frequency Electronics: Circuits and Applications.* Cambridge, UK: Cambridge University Press, 1996, pp. 32–40.

[146] E. R. Davies, *Electronics, Noise and Signal Recovery*, ser. Microelectronics and Signal Processing. London, UK: Academic Press, 1993, ch. 15, pp. 211–217.

[147] P. A. Petcher and S. Dixon, "Non-contact position tracking using phase correlation of sample images for NDT applications," *Insight*, vol. 53, no. 4, pp. 185–191, Apr. 2011.

[148] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[149] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Applied Optics*, vol. 23, no. 6, pp. 812–816, Mar. 1984.

[150] D. Psaltis and D. Casasent, "Deformation invariant optical processors using coordinate transformations," *Applied Optics*, vol. 16, no. 8, pp. 2288–2292, Aug. 1977.

[151] Y. N. Hsu, H. H. Arsenault, and G. April, "Rotation-invariant digital pattern recognition using circular harmonic expansion," *Applied Optics*, vol. 21, no. 22, pp. 4012–4015, Nov. 1982.

[152] M. P. Levesque and H. H. Arsenault, "Rotation-invariant pattern recognition using the phase of circular harmonic filter correlations," *Optics Communications*, vol. 58, no. 3, pp. 161–166, June 1986.

[153] D. M. Tsai and C. T. Lin, "Fast normalized cross correlation for defect detection," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2625–2631, Nov. 2003.

[154] A. J. H. Hii, C. E. Hann, J. G. Chase, and E. E. W. V. Houten, "Fast normalized cross correlation for motion tracking using basis functions," *Computer Methods and Programs in Biomedicine*, vol. 82, no. 2, pp. 144–156, May 2006.

[155] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[156] S. L. Marple, "Estimating group delay and phase delay via discrete-time analytic cross-correlation," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2604–2607, Sep. 1999.

[157] A. J. Jerri, "The Shannon sampling theorem - its various extensions and applications: a tutorial review," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.

[158] S. F. Burch, "Digital enhancement of video images for NDT," *NDT International*, vol. 20, no. 1, pp. 51–56, Feb. 1987.

[159] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, Mar. 2002.

[160] L. Leclerc, Y. Sheng, and H. H. Arsenault, "Rotation invariant phase-only and binary phase-only correlation," *Applied Optics*, vol. 28, no. 6, pp. 1251–1256, Mar. 1989.

[161] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, no. 2, pp. 156–158, Jan. 2008.

[162] R. Soummer, L. Pueyo, A. Sivaramakrishnan, and R. J. Vanderbei, "Fast computation of Lyot-style coronagraph propagation," *Optics Express*, vol. 15, no. 24, pp. 15 935–15 951, Nov. 2007.

[163] P. A. Petcher and S. Dixon, "Wideband ultrasonic time of flight diffraction combining B-scans and cross-sectional imaging," in *Review of Progress in Quantitative Nondestructive Evaluation: Proceedings of the 35th Annual Review of Progress in Quantitative Nondestructive Evaluation*, ser. AIP Conference Proceedings, vol. 1096. Chicago (Illinois), USA: AIP, 2009, pp. 650–657.

[164] X. Jian, S. Dixon, I. Baillie, R. S. Edwards, J. Morrison, and Y. Fan, "Shear wave generation using a spiral electromagnetic acoustic transducer," *Applied Physics Letters*, vol. 89, no. 24, p. 244106, Dec. 2006.

[165] P. A. Petcher and S. Dixon, "A modified Hough transform for removal of direct and reflected surface waves from B-scans," *NDT & E International*, vol. 44, no. 2, pp. 139–144, Mar. 2011.

[166] P. V. C. Hough, "Method and means for recognizing complex patterns," US Patent 3 069 654, Dec., 1962.

[167] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[168] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*, 2nd ed., ser. Signal Processing and its Applications. London, UK: Academic Press, 1997.

[169] F. O'Gorman and M. B. Clowes, "Finding picture edges through collinearity of feature points," *IEEE Transactions on Computers*, vol. 25, no. 4, pp. 449–456, Apr. 1976.

[170] S. I. Wright and B. L. Adams, "Automatic analysis of electron backscatter diffraction patterns," *Metallurgical Transactions A*, vol. 23, no. 3, pp. 759–767, Mar. 1992.

[171] K. Kunze, S. I. Wright, B. L. Adams, and D. J. Dingley, "Advances in automatic EBSP single orientation measurements," *Textures and Microstructures*, vol. 20, pp. 41–54, 1993.

[172] N. C. K. Lassen, "Automatic localisation of electron backscattering pattern bands from Hough transform," *Materials Science and Technology*, vol. 12, no. 10, pp. 837–843, Oct. 1996.

[173] N. C. Fitton and S. J. D. Cox, "Optimising the application of the Hough transform for automatic feature extraction from geoscientific images," *Computers & Geosciences*, vol. 24, no. 10, pp. 933–951, Dec. 1998.

[174] D. Chang and S. Hashimoto, "An inverse voting algorithm for Hough transform," in *Proceedings of the International Conference on Image Processing*, vol. 1. Los Alamitos (California), USA: IEEE Computer Society Press, Nov. 1994, pp. 223–227.

[175] M. Atiquzzaman and M. W. Akhtar, "Complete line segment description using the Hough transform," *Image and Vision Computing*, vol. 12, no. 5, pp. 267–273, June 1994.

[176] M. Atiquzzaman and M. W. Akhtar, "A robust Hough transform technique for complete line segment description," *Real-Time Imaging*, vol. 1, no. 6, pp. 419–426, Dec. 1995.

[177] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[178] J. Illingworth and J. Kittler, "A survey of the Hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87–116, Oct. 1988.

[179] S. A. Dudani and A. L. Luk, "Locating straight-line edge segments on outdoor scenes," *Pattern Recognition*, vol. 10, no. 3, pp. 145–157, 1978.

[180] R. S. Wallace, "A modified Hough transform for lines," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* San Francisco (California), USA: IEEE Computer Society Press, June 1985, pp. 665–667.

[181] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic Hough transform," *Pattern Recognition*, vol. 24, no. 4, pp. 303–316, 1991.

[182] P. A. Petcher and S. Dixon, "Parabola detection using matched filtering for ultrasound B-scans," *Ultrasonics*, vol. 52, no. 1, pp. 138–144, Jan. 2012.

[183] L. Capineri, P. Grande, and J. A. G. Temple, "Advanced image-processing technique for real-time interpretation of ground-penetrating radar images," *International Journal of Imaging Systems and Technology*, vol. 9, no. 1, pp. 51–59, 1998.

[184] M. K. Kuo, T. R. Lin, P. L. Liu, and T. T. Wu, "Locating the crack tip of a surface-breaking crack Part I. line crack," *Ultrasonics*, vol. 36, no. 7, pp. 803–811, May 1998.

[185] P. D. Wilcox, C. Holmes, and B. W. Drinkwater, "Advanced reflector characterisation with ultrasonic phased arrays in NDE applications," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 8, pp. 1541–1550, Aug. 2007.

[186] M. L. Boas, *Mathematical Methods in the Physical Sciences*, 2nd ed. New York, USA: John Wiley & Sons, 1983.

[187] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics and Engineering*, 2nd ed. Cambridge, UK: Cambridge University Press, 2002.

[188] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt, and J. M. Boone, *The Essential Physics of Medical Imaging*, 2nd ed. Philadelphia, USA: Lippincott Williams & Wilkins, 2002.

[189] D. York, N. M. Evensen, M. L. Martinez, and J. D. B. Delgado, "Unified equations for the slope, intercept, and standard errors of the best straight line," *American Journal of Physics*, vol. 72, no. 3, pp. 367–375, Mar. 2004.

[190] I. G. Hughes and T. P. A. Hase, *Measurements and Their Uncertainties.* Oxford, UK: Oxford University Press, 2010.